Docket No. DX0903K1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	
BAZAN *et al.*	Examiner: Chernyshev, Olga N
For: **MAMMALIAN CYTOKINES; RELATED REAGENTS AND METHODS**	Group Art Unit: 1649
Application No.: 09/963,347	Date: June 19, 2007
Filed: 09/25/2001	

Schering-Plough Corporation
Kenilworth, New Jersey 07033

COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, VA 22313-1450
Mail Stop: Amendment

DECLARATION OF ANDREJ SALI, PH.D. UNDER 37 C.F.R. § 1.132

I, Andrej Sali, Ph.D., hereby declare that:

1. I am a citizen of the United States and reside in San Francisco, California. My c.v. is attached as Exhibit A.

2. I have been retained by Schering-Plough to provide my opinion on certain issues in the proceedings involving U.S. Patent Application No. 09/963,347 ("the '347 application").

3. I have reviewed the following documents in the course of preparing this declaration:

   i. The '347 application

   ii. U.S. Provisional Application No. 60/101,318 ("the '318 provisional application")

   iii. Cosenza *et al.*, "Disulfide Bond Assignment in Human Interleukin-7 by Matrix-assisted Laser Desorption/Ionization Mass Spectroscopy and Site-directed Cysteine to Serine Mutational Analysis," J. BIOL. CHEM. 272(52): 32995-33000 (1997)

iv. Kroemer *et al.*, "Prediction of the three-dimensional structure of human interleukin-7 by homology modeling," PROT. ENG. 9(6): 493-498 (1996)

v. Pearson, "Effective protein sequence comparison," METHODS ENZYMOL. 266: 227-258 (1996)

vi. Mott and Campbell, "Four-helix bundle growth factors and their receptors: protein-protein interactions," CURR. OPIN. STRUCT. BIOL. 5(1): 114-121 (1995)

vii. Kruse *et al.*, "Two distinct functional sites of human interleukin 4 are identified by variants impaired in either receptor binding or receptor activation," EMBO J. 12(13): 5121-5129 (1993)

viii. Park *et al.*, "Intermediate sequences increase the detection of homology between sequences," J. MOL. BIOL. 273(1): 349-354 (1997)

ix. Rost, "PHD: predicting one-dimensional protein structure by profile-based neural networks," METHODS ENZYMOL. 266: 525-539 (1996)

4. I have worked in the field of bioinformatics since 1987 and have extensive experience in the fields of bioinformatics, computational structural biology, and predicting the structure and function of proteins in general. *See* Exh. A. I believe I am well-qualified to express an opinion regarding what the '318 provisional application would have conveyed to a person of skill in the art reading that application at the time it was filed.

5. I was asked to review the '318 provisional application to determine if a person of skill in the art at the time that the application was filed would have considered the statement relating to the ability of IL-B50 to stimulate or inhibit T cells and B cells to be credible in view of the disclosed similarities between IL-B50 and IL-7.

6. In order to determine whether or not the disclosed sequence similarity between IL-B50 and IL-7 is statistically significant, I used the program PRSS3 from the package FASTA; default values for all program options were used unless stated otherwise. *See, e.g.* Pearson, METHODS ENZYMOL. 266: 227-258 (1996); http://fasta.bioch.virginia.edu. PRSS3 was well known as of September of 1998 and was commonly used by those skilled in the art to evaluate the statistical significance of a similarity score between two given protein sequences.

7. PRSS3 first calculates an optimal local alignment and the corresponding "observed" similarity score using the Smith-Waterman algorithm for dynamic programming. It then calculates the distribution of the similarity scores for many pairs of original sequences whose

residues have been shuffled to make them "random." Next, an extreme value distribution is fit to this "random" distribution and the fitted parameters are used to estimate the probability (the p-value) that by chance alone the original sequences have a similarity score at least as high as the observed similarity score. In addition, the p-value is presented in a modified form as an E-value, which gives the number of matches with a score at least as high as the observed similarity score that would have arisen by chance when one of the original sequences is compared against a database containing a specified number of unrelated sequences. It is generally accepted that E-values smaller than 0.02 essentially prove homology. *See, e.g.* Pearson, METHODS ENZYMOL. 266: 227-258, 256 (1996).

8. Using the PRSS3 program, I found that the sequence similarity between IL-7 and IL-B50 is statistically significant when the comparison is made using an appropriately chosen database of human proteins known in 1998. In 1998, the SwissProt database contained approximately 60,000 proteins. Currently, SwissProt contains 270,778 proteins, of which 16,619 (or 6.1%) are human (6/15/07). *See* http://ca.expasy.org/sprot/relnotes/relstat. Assuming the proportion of the human sequences in SwissProt in 1998 was the same as today (*i.e.*, 6.1%), I estimated the number of human proteins in SwissProt in 1998 to be approximately 3,600. This is a reasonable estimate, considering the human genome sequencing efforts did not start producing human protein sequences in earnest until 1999.

9. The E-values (a parameter describing the number of hits that one can "expect" to see by chance when searching a database of a particular size) between IL-B50 and human and murine IL-7 when searching only human proteins as of September 1998 are approximately 0.16 and 0.01, respectively. The latter E-value is statistically significant. While the E-value of 0.16 between the human IL-B50 and human IL-7 is not considered to guarantee homology on its own, it still indicates homology per generally accepted standards. *See, e.g.* Pearson, METHODS ENZYMOL. 266: 227-258, 256 (1996). Moreover, because there is unequivocal homology between human IL-7 and murine IL-7 as demonstrated by the statistically significant sequence similarity score of 0.01, there must also be homology between the human IL-B50 and human IL-7. This kind of transitive logic is generally accepted in detecting homology based on an intermediate sequence search approach. *See, e.g.* Park *et al.*, J. MOL. BIOL. 273: 349-354 (1997). I note that even if there were twice as many known human protein sequences as estimated here, that fact would not have changed the statistical significance of the observed similarity scores. Therefore, my conclusions are not dependent on a precise estimate of the number of human protein sequences known in 1998.

10. The version of PRSS3 that is currently available is not exactly that same as the version available in 1998, however they both use the same algorithms for calculating the optimal alignment and the corresponding

E-value. Therefore, to the best of my knowledge, the E-values calculated with the version of the PRSS3 program that existed in 1998 would be essentially the same as the E-values I calculated using the currently available program.

11. I additionally compared the predicted secondary structures of IL-7 and IL-B50. The secondary structures were predicted by the program PhD, which was available in 1998. *See* Rost, METHODS ENZYMOL. 266: 525-539 (1996). The predicted secondary structures of IL-7 and IL-B50 suggest that the two proteins indeed have the four helices typical of the cytokine 4-helix bundle. This finding is substantiated by independent predictions for IL-7 made by others and published before 1998. *See* Kroemer *et al.*, PROT. ENG. 9: 493-498 (1996). Further, the IL-7 regions found (by site-directed mutagenesis) to be important for binding other proteins, including the receptor, are also the regions where the similarity to IL-B50 is highest, including helices A and D as well as loops A-B and C-D. *See, e.g.* Mott and Campbell, CURR. OPIN. STRUCT. BIOL. 5(1): 114-121 (1995); Kruse *et al.*, EMBO J. 12(13): 5121-5129 (1993); '318 provisional application at p. 11.

12. Further, IL-B50 and IL-7 share conserved short sequence patterns indicating PKC phosphorylation, Asn glycosylation, and myristilation sites. The proteins also share one approximately conserved pattern indicating protein kinase CK2 phosphorylation. These functional sequence motifs were well characterized within the art prior to September 1998.

13. Based on: (i) the statistically significant similarity of the IL-7 and IL-B50 sequences, (ii) the similarity between the predicted secondary structures of IL-7 and IL-B50, (iii) the higher similarity in known functionally important regions compared to the rest of the sequence, and (iv) the conserved short functional sequence motifs, if I were reading the '318 provisional application at the time that it was filed, I would have believed that IL-B50 and IL-7 have similar sequences and biological functions. Accordingly, I would have found the utility asserted in the '318 provisional application to be credible.

I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such will false statements may jeopardize the validity of the patent application or any patent issuing thereon.

Andrej Šali — 6/19/07

Andrej Sali, Ph.D. — Date

EXHIBIT A

Curriculum vitæ

Name: **Andrej Sali**

Position: Professor, Step V
Department of Biopharmaceutical Sciences
Department of Pharmaceutical Chemistry
California Institute for Quantitative Biomedical Research (QB3)
School of Pharmacy

Integrative Program in Quantitative Biology:
Bioinformatics and Medical Informatics Graduate Program
Biophysics Graduate Program
Graduate Program in Complex Biological Systems
Chemistry and Chemical Biology Graduate Program
Pharmaceutical Sciences and Pharmacogenomics Graduate Program

Address: Byers Hall, Suite 503B, MC 2552
University of California at San Francisco
1700 4th Street
San Francisco, CA 94158-2330

Voice: +1-415-514 4227
Fax: +1-415-514 4231
email: sali@salilab.org
web: http://salilab.org

EDUCATION

1983-1987	University of Ljubljana, Slovenia	BSc	Chemistry
1987-1991	University of London, UK	PhD	Molecular Biophysics

PRINCIPAL POSITIONS HELD

1987-1991 PhD student. Department of Crystallography, Birkbeck College, and Imperial Cancer Research Fund, London, UK (Mentor: Prof. Thomas L. Blundell, FRS).

1991-1994 Postdoctoral Fellow at Department of Chemistry, Harvard University, Cambridge, USA (Mentor: Prof. Martin Karplus).

1995-2000	Assistant Professor, Head of Laboratory. The Rockefeller University.
2000-2003	Associate Professor, Head of Laboratory. The Rockefeller University.
2003-	Professor in Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research. University of California at San Francisco.

OTHER POSITIONS HELD CONCURRENTLY

2003-	Vice Chair of Computational Biology in Dept. of Biopharmaceutical Sciences, University of California at San Francisco.

HONORS, AWARDS, AND FELLOWSHIPS

1984-1987	Undergraduate scholarship from J. Stefan Institute, Ljubljana, Slovenia.
1985	British Council Visiting Student at Birkbeck College, London, UK.
1987	British Council Visiting Student at Birkbeck College, London, UK.
1987-1990	Overseas Research Students Award from the Committee of Principals and Vice Chancelors, England.
1987-1988	Scholarship for graduate studies from the Research Council of Slovenia.
1989-1990	Merck Sharp & Dohm academic scholarship.
1991-1994	Jane Coffin Childs Memorial Fund for Medical Research Postdoctoral Fellow.
1996-1999	Alexandrine and Alexander L. Sinsheimer Scholar.
1998-2000	Alfred P. Sloan Research Fellow.
2000-2003	Irma T. Hirschl Career Award Scientist.

KEYWORDS/AREAS OF INTEREST:

We employ the laws of physics and the rules of evolution to develop and apply methods for:

- predicting the structures of proteins;
- determining the structures of macromolecular assemblies;
- annotating the functions of proteins and their assemblies using their structures.

SOFTWARE

1993	MODELLER, a program for comparative protein structure modeling by satisfaction of spatial restraints; licensed to Accelrys Inc. since 1994.

2000 MODPIPE, a program for large-scale comparative protein structure modeling; licensed to Accelrys Inc. and Structural Genomix Inc. (2000-2004)

2000 MODBASE, a comprehensive database of comparative protein structure models; licensed to Structural Genomix Inc. (2000-2004)

PROFESSIONAL ACTIVITIES

Member of professional societies:

- Protein Society.
- AAAS.

Organization of meetings:

- Program Committee, Georgia Tech Intl. Conference on Bioinformatics, Atlanta GA. 1999-2001.
- Program Committee, Math/Chem/Comp 2001, Dubrovnik, Croatia. June 25-30, 2001.
- Program Committee, Protein Society 19th Symposium in Boston, July 2005.
- Program Committee, Protein Structure Modeling Workshop, Rutgers University, NJ, November 2005.
- Program Committee, American Society for Biochemistry and Molecular Biology, April 2006.
- Organizer, World Molecular Engineering Network, annual TSRI & UCSF meeting in San Jose del Cabo, Mexico, 2006-.
- Program Committee, 4th Conference on Modeling of Protein Interactions – MPI 2007

Service to professional publications:

- Editor, Structure.
- Member, Editorial Board, PLOS Computational Biology; Journal of Computer Aided Molecular Design; Molecular and Cellular Proteomics; Protein Engineering, Design, and Selection.
- Section Head for the Structural Genomics section on BioMed Central.
- Reviewer for Nature, Science, Cell, Proc.Natl.Acad.Sci.USA, Nat. Struct. Mol. Biol, Nat.Gen., Nat.Biotech., Structure, J. Mol.Biol., Proteins, Prot.Eng., Prot.Sci., Bioinformatics, Nucl.Acids Res., J.Biol.Chem., BMC Structural Biology, Genome Biology, FEBS Letters, J.Com.Aid.Mol.Des., Biophysical Journal, Biochemical Journal, PLoS Biology, and PLoS Computational Biology.

Government and other professional service:

- Executive Committee of the Protein Society (2005-).
- Advisory Committee of the Protein Data Bank (2005-)
- NIH/GM MSF-B Study Section (successor of BBCA) (2004-).

- Reviewer for NIH, NSF, DOE, Burroughs Wellcome Fund, European Community, and BSF grant proposals.

INVITED PRESENTATIONS (since 1999)

1. Structural Genomics Targets Workshop, NIH, Washington DC, USA. February 11, 1999.
2. Advances & Opportunities at the Biology/Math/Computational/Physical Sciences Interface, Rutgers University, New Brunswick, New Jersey, USA. March 6, 1999.
3. Mount Sinai School of Medicine, New York, New York, USA. March 19, 1999.
4. New York Structural Biology Group, New York Academy of Sciences, New York, New York, USA. March 24, 1999.
5. Columbia University, New York, New York, USA, 1999.
6. Chemistry Dept., New York University, New York, New York, USA, 1999.
7. CERCA CADD Symposium, Montreal, Canada. April 13, 1999.
8. Protein Sequence Structure Function Meeting, UCSF, San Francisco, California, USA. April 23, 1999.
9. Data Mining in Crystallography, Erice, Italy. May 15, 1999.
10. Structural Biology Net, Tallberg, Sweden. June, 1999.
11. The Scripps Institute, La Jolla, California, USA. August 13, 1999.
12. Frontiers in Structural Biology, Indian Institute of Science, Bangalore, India. August 27, 1999.
13. Mathematical Problems in the Molecular Sciences, Courant Institute, New York, New York, USA. October 10, 1999.
14. City College of New York, New York, USA. October 20, 1999.
15. Agouron Pharmaceuticals, San Diego, California, USA. October 28, 1999.

16. Second International Georgia Tech Conference in Bioinformatics, Atlanta, Georgia, USA. November 12, 1999.

17. Structural Genomics Conference, ANL, Chicago, Illinois, USA. November 16, 1999.

18. Structural Genomics and the Pharmaceutical Industry, Princeton, New Jersey, USA. November 18, 1999.

19. BRI, Montreal, Canada. November 24, 1999.

20. University of Toronto, Toronto, Canada. December 9, 1999.

21. Quantitative Challenges in the Post Genomic Sequence Era, La Jolla Interfaces in Science, San Diego, California, USA. January 12, 2000.

22. Japan Biophysical Society Meeting, Tokyo, Japan. January 17, 2000.

23. UCSD, Dept of Physics, San Diego, California, USA. January 19, 2000.

24. UCSF, San Francisco, California, USA. January 20, 2000.

25. Center for Physics and Biology, Rockefeller University,New York, New York, USA. January 24, 2000.

26. Biological Chemistry Seminar Series, University of Penn, Philadelphia, Pennsylvania, USA. February 17, 2000.

27. ABRF 2000 ``From Singular to Global Analyses of Biological Systems", Bellevue, Washington, USA. February 22, 2000.

28. AAAS conference, Washington DC, USA. March 20, 2000.

29. Keystone Symposium on Macromolecular Assemblies at Work: Application of Physics, Chemistry, and Mathematics to Biology, Durango, Colorado, USA. March 25, 2000.

30. Bio2000, Boston, Massachusetts, USA. March 28, 2000.

31. Bioinformatics 2000, Elsinore, Denmark. April 28, 2000.

32. Computational Challenges of the Post Genomic Age, SUN, San Francisco, California, USA.

May 12, 2000.

33. Biopolymers Gordon Conference, Newport, Rhode Island, USA. June 18-22, 2000.

34. 2000 FASEB Summer Research Conference on Protein Folding in the Cell, Saxton River, Vermont, USA. July 22-27, 2000.

35. Monsanto/Pharmacia lectureship series, Univ. of Saint Louis, Missouri, USA. September 28, 2000.

36. Workshop on Structural Genomics. NIGMS, Washington DC, USA. October 23, 2000.

37. Genomics and Bioinformatics, UMD, New Brunswick, New Jersey, USA. November 2, 2000.

38. University of Minnesota, Minneapolis, Minnesota, USA. November 27, 2000.

39. Oncogenomics: Dissecting Cancer Through Genome Research, Nature Genetics, Tuscon, Arizona, USA. January 25-27, 2001.

40. UAB, Birmingham, Alabama, USA. February 26, 2001.

41. Structural Genomics and Bioinformatics, Instituto Juan March, Madrid, Spain. March 12-14, 2001.

42. Bard College, New York, USA. April 18, 2001.

43. Physics/Chemistry. CSUN, Northridge, California, USA. May 2, 2001.

44. Annual meeting of the Canadian Society for Biochemistry and Molecular and Cellular Biology (CSBMCB), Toronto, Canada. May 31-June 3, 2001.

45. Math/Chem/Comp 2001, Dubrovnik, Croatia. June 25-30, 2001.

46. 4th International Conference on Biological Physics, ICBP2001, Kyoto, Japan. July 30-August 3, 2001.

47. ACS Meeting, Chicago, Illinois, USA. August 26-30, 2001.

48. 4th International Conference on Molecular Structural Biology, ICMSB2001, Vienna, Austria.

September 5-9, 2001.

49. University of Maryland, Maryland, USA. October 2, 2001.

50. Columbia University, New York, New York, USA. October 15, 2001.

51. Genomics & Proteomics meeting, Barcelona, Spain. October 19, 2001.

52. Mast Cell Workshop, Bethesda, Maryland, USA. November 26-30, 2001.

53. University of Zuerich, Zuerich, Switzerland. December 5, 2001.

54. Bioinformatics & Proteomics: From Sequence to Function, Lausanne, Switzerland. December 6, 2001.

55. Genomics Seminar Series, Skirball Institute, New York, New York, USA. February 6, 2002.

56. Mining the Human Genome for New Drug Discovery - New Ways of Handling Orphan Targets. NYAS, New York, New York, USA. February 26, 2002.

57. Biological Processes for New and Innovative Engineering Systems and Applications, ARO workshop, Research Triangle Park, North Carolina, USA. February 26-27, 2002.

58. New York City Blood Centre, New York, New York, USA. March 7, 2002.

59. Proteomics-The New Frontiers: Discovery, Separation, Prediction & Modeling, University of Delaware, Newark, Delaware, USA. March 14-15, 2002.

60. Harvard University, Cambridge, Massachusetts, USA, March 28, 2002.

61. Molecular Cell Biology and Biochemistry Seminar Series, Virginia Tech, Blacksburg, Virginia, USA. April 5, 2002.

62. UCSF, San Francisco, California, USA, April 15, 2002.

63. A Workshop on large biological structures, Asilomar, California, USA. April 20-22, 2002.

64. SCBMB Program, Baylor College of Medicine, Houston, Texas, USA. May 15, 2002.

65. University of Barcelona, Barcelona, Spain. May 21, 2002.

66. Genomics & Proteomics meeting, Barcelona, Spain. May 22, 2002.

67. Samuel Lunenfeld Research Institute, Toronto, Canada. May 29, 2002.

68. 50th ASMS conference American Society of Mass Spectrometry, Orlando, Florida, USA. June 2-6, 2002.

69. The 5th Summer Session of the New York Structural Biology Discussion Group , Cold Spring Harbor Laboratory, New York, USA. June 26, 2002.

70. Berkeley-Stanford summer school for protein crystallography, SSRL, Stanford, California, USA. July 8-12, 2002.

71. Gordon Conference on Diffraction Methods in Structural Biology, Connecticut College, New London, Connecticut, USA. July 14-19, 2002.

72. The 19th Congress and General Assembly of the International Union of Crystallography IUCR, Geneva, Switzerland. August 6-15, 2002.

73. NYCBS New York Computational Biology Society seminar, NAS, New York, New York, USA. September 18, 2002.

74. Center for Biological Modeling, Michigan State University, East Lansing, Michigan, USA. September 27, 2002.

75. Bioinformatics seminar, Texas A&M University, Tamu, TX, USA. November 7, 2002.

76. Genomics and Bioinformatics Center Inaugural Symposium, Pontificia Universidad Catolica, Santiago, Chile. November 18-20, 2002.

77. Structure and Function of the Proteome, Argonne National Laboratory, Argonne, Illinois, USA. November 23-24, 2002.

78. Keystone Symposium in Proteomics: Technologies and Applications, Keystone Resort in Keystone, Colorado, USA. March 25-30, 2003.

79. NCRR sponsored Workshop on Structural Proteomics of Complexes, Bethesda, Maryland, USA. April 7-8, 2003.

80. American Society for Biochemistry and Molecular Biology meeting, San Diego, California, USA. April 11-15, 2003.

81. St. Jude Children's Research Hospital, Memphis, TN, USA. April 22, 2003.

82. Genentech, Inc., South San Francisco, CA, USA. April 29, 2003.

83. Fourteenth Annual World Molecular Engineering Network (WMEN) Conference, San Jose del Cabo, Baja California Sur, Mexico. May 4-8, 2003.

84. Structure and Function of Proteome, SBC, Argonne National Laboratory, Argonne, Illinois, USA. Spring, 2003.

85. "Frontiers of Bioinformatics" symposium, Center of Excellence in Bioinformatics, University at Buffalo, Buffalo, New York, USA. June 6-8, 2003.

86. IBM Thomas J. Watson Research Center, New York, New York, USA. June 11, 2003.

87. 2003 Gordon Research Conference on 3D Electron Microscopy of Macromolecules, Colby Sawyer College, New London, New Hampshire, USA. June 22-26, 2003.

88. PSI workshop ond data management, NIH Campus, Bethesda, MD, USA. July 10-11, 2003.

89. GTL and Beyond: Data and Computational Needs Workshop, San Francisco, CA, USA. September 10-11, 2003.

90. 2003 Pharmaceutical Sciences and Pharmacogenomics Retreat, Marshall, CA, USA. September 11-13, 2003.

91. Structure and Chemistry Seminar at Scripps, San Diego, CA, USA. September 18, 2003.

92. 5th Meeting of the Slovenian Biochemical Society, Ljubljana, Slovenia. September 24-28, 2003.

93. Seminar at Northeastern University, Boston, MA, USA. October 6, 2003

94. Workshop on Visualization of Biological Complexes, Four Points Sheraton Hotel, Emeryville, San Francisco Bay Bridge, CA, USA. October 11-12, 2003

95. Seminar at PARC, Palo Alto, CA, USA. October 15, 2003.

96. NIGMS Homology Modeling Workshop, Bethesda, MD, USA. October 21-22, 2003.

97. Seminar at Purdue University, West Lafayette, IN, USA. October 24-25, 2003.

98. PSI Target Selection Workshop, Bethesda, MD, USA. November 13-14, 2003.

99. Biophysics/CCB Retreat, Asilomar Conference Center, Pacific Grove, CA, USA. December 7-9, 2003.

100. University of Cologne, Cologne, Germany. January 19, 2004.

101. Ringberg meeting, Schloss Ringberg, Germany. January 21-23, 2004.

102. Licensing Executives Society meeting, San Francisco, CA, USA. February 12, 2004.

103. The Structural, Functional and Evolutionary Gordon Conference, Four Points Sheraton Harbortown, Ventura, CA, USA. February 15-20, 2004.

104. Seminar at Berkeley, Berkeley, CA, USA. March 8, 2004.

105. Seminar at UCSC, Santa Cruz, CA, USA. March 11, 2004.

106. Workshop on Structure Determination of Macromolecular Machines and Assemblies by Hybrid Methods, Granlibakken/Lake Tahoe Conference Center, CA, USA. March 17-20, 2004.

107. Workshop on Structural Determination of Environmentally Responsive Gene (ERG) Products for Diagnostics & Drug Discovery (NIEHS/DERT), Snowbird Resort, Snowbird, Utah, USA April 12-13, 2004.

108. 2004 Keystone Symposium on Structural Genomics, Snowbird Resort, Snowbird, Utah, USA April 13-19, 2004.

109. BayGenomics PGA, San Francisco, CA, USA. April 27, 2004.

110. Fourteenth Annual World Molecular Engineering Network (WMEN) Conference (2004), San Jose del Cabo, Baja California Sur, Mexico. May 2-6, 2004.

111. Gladstone Scientific Retreat, Asilomar in Monterey County, CA, USA. May 18-20. 2004.

112. Seminar at Caltech, Pasadena, CA, USA. October 12, 2004.

113. EMBO conference on Structures in Biology, EMBL, Heidelberg, Germany. November 10-13, 2004.

114. ICSG 2004 Meeting, Washington, DC, USA. November 17-24, 2004.

115. Workshop of the Center of Protein Folding Machinery, Stanford University, CA, USA. December 4-5, 2004.

116. Biological and Medical Informatics/Biophysics/Chemistry and Chemical Biology graduate groups retreat, Asilomar Conference Center, Pacific Grove, CA, USA. December 5-7, 2004.

117. The 7th World Congress of the World Association of Theoretically Oriented Chemists (WATOC), Capetown, South Africa. January 16-21, 2005.

118. ABRF meeting, Biomolecular Technologies:Discovery to Hypothesis, Savanah, Georgia, USA. February 5-8, 2005.

119. Speaker at Frontiers in Computational Biophysics Symposium, NIH campus in Bethesda, MD, USA. April 29-30, 2005.

120. NIH Symposium on Structural Analysis of Large Assemblies: Sizing up the Challenges, NIH campus in Bethesda, MD, USA. June 2-3, 2005.

121. Speaker at SRI's Computational Biology series, SRI International, Menlo Park, CA, USA. June 29, 2005.

122. 19th Annual Symposium of the Protein Society, Boston, MA, USA. July 30 - August 3, 2005.

123. GRC 2005 Computer-aided design meeting, Tilton School, NH, USA. July 31 - August 5, 2005.

124. Keynote Speaker in XX IUCr Congress in Firenze, Italy. August 23-31, 2005.

125. Seminar at the Biochemical and Biophysical Methods Course Fall 2005, The Rockefeller University, New York, NY, USA. October 11, 2005.

126. Seminar at the Novartis Institutes for BioMedical Research, Cambridge, MA, USA.

November 9, 2005.

127. Workshop on Biological Macromolecular Structure Models, The State University of New Jersey, Piscataway, NJ, USA. November 19-20, 2005.

128. Speaker at the International Workshop M2CELL, The Royal Abbey of Fontevraud, Paris, France. December 4-6, 2005.

129. Organizer and Speaker of the Theme "Macromolecular Structure and Dynamics" with 4 Symposia. ASBMB 2006 meeting, San Francisco, CA, USA. April 1-5, 2006.

130. Organizer and Speaker of the World Molecular Engineering Network Conferenc. Cabo San Luca, Mexico. April 30-May 2nd, 2006.

131. Plenary Speaker at the 11th Sympsosium on Recent Advances in Biophysics, National Taiwan University, Taipei, Taiwan. May 23-26, 2006

132. Seminar at UC Davis, CA, USA. June 1, 2006.

133. 2006 Keystone Symposium on Multi-Protein Complexes Involved in Cell Regulation, St. John's College, Cambridge, UK. August 18-23, 2006.

134. Symposium at Wyeth Research, Cambridge, MA, USA. October 16, 2006.

135. Seminar at the University of Massachusetts, Dept. of Biochemistry and Molecular Pharmacology, Worcester, MA, USA. October 18, 2006.

136. Seminar at the Fifth Annual Systems Biology Course at the Institute for Systems Biology, Seattle, WA, USA. November 9, 2006.

137. Seminar at UC Merced Center for Computational Biology, Merced, CA, USA. November 30, 2006.

138. Speaker at the TDR/WHO Drug Target Selection Meeting in Seattle, OR, USA. December 1st, 2006.

139. Speaker at the Biological and Medical Informatics/Biophysics/Chemistry and Chemical Biology Graduate Groups Retreat, Monterey, CA, USA. December 3-5, 2006.

140. Speaker at the Collaborative Drug Discovery meeting, UCSF, San Francisco, USA. March 1st 2007.

141. Organizer and Speaker of the World Molecular Engineering Network Conference. Cabo San Lucas, Mexico. April 29-May 2nd, 2007.

142. Invited Speaker at the Protein Complexes and Protein Networks Symposium in Martinsried, Germany, May 21-22, 2007.

143. Speaker at The Protein Folding Center Annual Retreat, Stanford, CA, USA. May 27-29 2007.

144. Invited Speaker at the Center for Theoretical Biological Physics. UCSD,San Diego USA. June 1st, 2007.

145. Speaker and Organizer of the Modeling of Protein Interactions Meeting, MPI-2007, Lawrence, KA, USA, September 30-October 2, 2007.

146. Invited Speaker at the Keystone Symposium on Structural Genomics and Its Applications to Chemistry, Biology and Medicine, Steamboat Springs, Colorado, CO, USA, January 6-11 2008.

147. 40th Course: From Molecules to Medicines Integrating Crystallography in Drug Discovery, Erice, Italy. May 29 - June 8, 2008.

UNIVERSITY COMMITTEES

- Faculty steering committee for the shared computer cluster at QB3, Chair (2006)
- Basic Sciences Research Resources Oversight Committee, UCSF (2005-)
- Rock Hall Governance Committee, UCSF (2005-)
- QB3 Executive Committee (2004-)
- QB3 Governance/Community Committee, Chair (2005-)
- Mission Bay leadership committee, UCSF (2005-)
- BPS Internal Advisory Committee, UCSF (2003-)
- Biophysics Graduate Program Executive Committee, UCSF (2003-)
- Systems Biology Faculty Search Committee (2005-)
- MD/PhD Scientist at QB3 Search Committee, UCSF (2003-)
- Bioinformatics and Medical Informatics Graduate Program Executive and Admissions Committee, UCSF (2003-)
- Information Technology at School of Pharmacy Committee, UCSF (2003-)
- QB3 Building Committee, UCSF (2003-2005)
- Pharmaceutical Sciences and Pharmacogenomics Graduate Program Admissions Committee, UCSF (2003-2005)

- School of Pharmacy Strategic Planning Committee, UCSF (2005)
- Chancellor's Council committee, UCSF (2004)
- Academic Information Technology Coordinator search committee, UCSF (2004)
- Future of Computing at UCSF Committee (2003)
- Bioinformatics and Computational Biology Faculty Search Committee, Co-Chair, UCSF (2003-2004)
- Bioinformatics and Computational Biology Faculty Search Committee, UCSF (2002-2003)
- Ad Hoc Faculty reviews (~15 since 2003)
- Computer Security Committee, Chair, Rockefeller University (2000).
- Faculty Awards Nominations Committee, Rockefeller University (2000-2002)
- Bioinformatics Search Committee, Rockefeller University (2000-2002)
- Dean's Graduate Studies Admissions Committee, Rockefeller University (2000-2002)
- Chemistry Search Committee, Rockefeller University (1995-2000)

TEACHING

- Instructor (with Prof. Patsy Babbitt and Tanja Kortemme) for a yearly one semester graduate course "Bioinformatics" BMI-206 at UCSF, since 2004.
- Instructor (with Prof. Patsy Babbitt) for a yearly one semester SOP course "Bioinformatics" BPS-114 at UCSF, since 2005.
- Invited lecturing at several other courses and seminar series at UCSF (Bioengineering, Macromolecules).
- Chemistry and Chemical Biology Graduate Program Faculty member at UCSF, since 2004.
- Instructor for a one semester graduate course "Analysis and prediction of protein structures" at The Rockefeller University, in 1998, and 2000.
- Guest lecturer at courses and workshops at Rockefeller University (1995-2002), Weill Medical College of Cornell University (1999-2002), New York University (2000-2002), Crystallography School in Erice, Italy (1999), FEBS course in Barcelona, Spain (1990), and Stanford University (2002).

PREDOCTORAL MENTORING:

Current graduate students:

- Mark Peterson, BMI (shared with Prof. Patsy Babbitt) (6/1/2003)
- Michael Kim, BMI (6/1/2003)
- Ranyee Chiang, BMI (shared with Prof. Patsy Babbitt) (9/1/2002)

- Libusha Kelly, BMI (6/1/2003)
- David Eramian, Biophysics (8/1/2004)
- Keren Lasker (shared with Prof. Haim Wolfson, Tel-Aviv University) 8/1/2006)
- David Barkan, BMI (9/1/2006)

Past graduate students:

- Roberto Sanchez (8/20/1995-10/6/2000); Assistant Professor at Mount Mount Sinai School of Medicine, NY, since 10/9/2000
- Eric Feyfant (1999-2001; postdoctoral fellow 2/10/2001-11/1/2001), scientist at Genetics Institute Inc. (now Wyeth Inc.), Cambridge, MA, since 11/2/2001
- Bino John (8/1/2000-4/19/2003); Assistant Professor at The University of Pittsburgh School of Medicine, since 9/1/2005
- Nebojsa Mirkovic (4/1/1999-10/1/2003); postdoctoral fellow with Diana Murray at Weill Medical College of Cornell University, NY, since 2/1/2004
- Fred Davis, Biophysics (1/1/2003-3/1/2007)

POSTDOCTORAL MENTORING:

Current Postdoctoral Fellows:

- Ursula Pieper (2000)
- Narayanan Eswar (2000)
- Mallur S. Madhusudhan (3/1/2000)
- Frank Alber (04/15/2001)
- Dmitry Korkin (9/1/2002)
- Min-yi Shen (2/1/2003)
- Ben Webb (9/1/2003)
- Friedrich Foerster (7/1/2005)
- Hao Fan (9/1/2006)
- Javier Velazquez (11/13/2006)
- Daniel Russel (02/01/2007)

Past Postdoctoral Fellows:

- Ilya Vakser (1995-1997), Assistant Professor at University of South Carolina at Charleston until 2002; currently Associate Professor at SUNY Stony Brook
- Andras Fiser (9/15/1997-12/31/2002), Assistant Professor at Albert Einstein College of Medicine, Bronx, NY, since 1/2/2003
- Ash Stuart (7/1/1999-3/2003), Assistant Professor at Ramapo College, Mahwah, NJ.
- Azat Badretdinov (1996-1999); senior scientific programmer at Accelrys Inc., San Diego,

since 1999

- Francisco Melo (9/1/1998-1/1/2001), Assistant Professor at Pontificia Universidad Catolica de Chile, since 1/1/2001
- Bozidar Yerkovich (2002-2003), Head of Structural Bioinformatics at Rosetta Inpharmatics Inc., Seattle, since 2003
- Valya Ilyin (2000-2002), Assistant Professor at Northeastern University, Boston, since 2002
- Marc A. Marti-Renom (2/1/1999-5/31/2006); Adjunct Assistant Professor at UCSF, since 2/1/2003; Assistant Professor at Prince Felipe Research Center, Valencia, Spain, since 7/1/2006
- Rachel Karchin (8/1/2003-8/31/06); Assistant Professor at Johns Hopkins University
- Damien Devos (11/15/2002-6/30/06); at EMBL, Heidelberg.
- Maya Topf (2/1/2003-10/30/06); Lecturer in Dept. of Crystallography, Birkbeck College, London, since 11/1/06.
- Niu Huang (12/1/2003)
- Andrea Rossi, Senior Scientist at Rinat Laboratories, Pfizer Inc (11/1/2000-11/2006)

FACULTY MENTORING

- Assigned faculty mentor for Tanja Kortemme (BPS)

DOCTORAL DISSERTATION AND ORAL QUALIFYING EXAMINATION COMMITTEES (since 2004)

2004	Tiba Ayunechi (Doctoral Dissertation)
2004	Barbara Novak, BMI (Orals)
2004	Alan Graves, Biophysics (Orals)
2004	Mike Kim, BMI (Thesis Committee)
2004	Alexandra Schnoes (Orals)
2005	Ben Sellers, Biophysics (Orals)
2005	Fred Davis, Biophysics (Thesis Committee)
2005	Nathan Salomonis, PSPG (Orals)
2005	Jerome Nilmeir, Biophysics (Orals)
2005	Libusha Kelly, BMI (Thesis Committee)
2005	Tuan Pham, BMI (Orals)
2005	Ranyee Chiang, BMI (Thesis Committee)
2005	Mark Peterson, BMI (Thesis Committee)
2006	Marco Sorani, BMI (Orals)

2006	David Lomelin, BMI (Orals)
2006	Nima Fayazmanesh, Biophysics (Orals)
2006	Arjun Narayanan, Biophysics (Orals)
2006	David Eramian, Biophysics (Thesis Committee)
2006	Veena Thomas, PSPG (Orals)
2006	Dale Webster, BMI (Orals)
2006	Holly Atkinson, BMI (Orals)
2006	Dan Mandel, BMI (Orals)
2006	Mike Keiser, BMI (Orals)
2007	Elisabeth Humphris (Orals)
2007	Colin A. Smith (Orals)
2007	Greg Friedland (Thesis Committee)

SUMMARY OF TEACHING HOURS:

2003-04: 50 hours of teaching.
Formal class or course teaching hours: 18 hours.
Informal teaching hours: 4 hours.
Mentoring hours: 500 hours.

2004-05: 100 hours of teaching.
Formal class or course teaching hours: 30 hours.
Informal teaching hours: 6 hours.
Mentoring hours: 500 hours.

2005-06: 100 hours of teaching.
Formal class or course teaching hours: 30 hours.
Informal teaching hours: 6 hours.
Mentoring hours: 500 hours.

2006-07: 100 hours of teaching.
Formal class or course teaching hours: 30 hours.
Informal teaching hours: 6 hours.
Mentoring hours: 500 hours.

2007-2008 : Anticipated total teaching load is expected to be similar to that in the previous year.

Since my arrival to UCSF in January 2003, I joined Prof. Patsy Babbit in leading the graduate and professional students' courses in Bioinformatics (BMI-206 and BPS-114). Patsy and I also give many of the lectures and supervise student activity associated with the course (student seminars, exercise sets). I also participate as a guest lecturer in a number of other courses, such as Macromolecules.

CONSULTING

- Founder and consultant, Prospect Genomics Inc., San Francisco, now Structural Genomix Inc., San Diego (1999 to 2004).
- Consultant, Accelrys Inc., San Diego (1994 to date).
- Consultant, Biogen Inc., Cambridge, MA, 2000-2002.
- Consultant, Millenium Inc., Cambridge, MA, 2001-2002.
- Consultant, Wyeth Inc., Princeton, NJ, 2006.

RESEARCH AWARDS AND GRANTS

07/01/96 – 06/30/01	Protein modeling by satisfaction of spatial restraints. NIH/NIGMS R29 GM 54762. $350,000 in direct costs over the course of the grant period.
09/01/96 – 08/31/98	Sinsheimer Scholar Award from Alexandrine and Alexander L. Sinsheimer Fund. $120,000.
10/01/96 – 09/30/98	Acquisition of a multiprocessor computer for computational physics and structural biology. NSF BIR-9601845. $289,000.
10/01/98 – 09/30/00	Alfred P. Sloan Fellowship. $40,000.
07/01/98 – 06/30/99	Gift. Molecular Simulations Inc. $8,500.
09/01/99 – 08/31/03	Disruption and expression of mast cell protease genes. NIH/NHLBI R01 HL63284-02 (R. Stevens, PI; A. Sali, consultant). $145,000.
07/01/00 – 06/30/03	Gift. Prospect Genomics Inc. $180,000.
01/01/00 – 12/31/04	Career Scientist Award from Irma T. Hirschl Monique Weill-Caulier Trust. $125,000.
01/01/00 – 12/31/02	Comparative annotation of eukaryotic genomes: From gene detection to protein structure modeling. Mathers Foundation Award. $270,000.
02/01/00 – 01/31/03	Target selection for the structural genomics of cancer. NIH/NCI R33 CA84699 (T. Gaasterland, PI; A. Sali, Co-PI). $270,000.
02/01/00 – 01/31/02	Database of comparative protein structure models for genomics. The Merck Genome Research Institute. $260,000.
09/30/00 – 08/31/05	Center for Structural genomics. NIH/NIGMS P50 GM62529 (S.K. Burley, PI; A. Sali, co-PI). $850,000.
06/01/01 – 03/31/04	Comprehensive map of cellular protein interactions. NIH/NCI R33 CA89810 (B. Chait, PI; A. Sali, consultant). $170,000.
07/01/01 – 06/30/10	Protein modeling by satisfaction of spatial restraints. NIH/NIGMS R01 GM 54762. $1,400,000.
07/03/01	Development of an integrated software environment for high-throughput

	structural biology and automated comparative protein structure modeling. Sun Academic Equipment Grant. EDUD-7824-020257-US, $294,185.
07/01/03 – 06/30/06	3D-reconstruction and identification of postsynaptic molecular complexes images by electron cryotomography. Human Frontier Science Program Research Grant. RGP67/2003. (W. Baumeister, PI; A. Sali, Co-Investigator). $225,000.
04/01/03 - 3/31/10	Pharmacogenetics of Membrane Transporters. U01 GM61390 (K. Giacomini, PI; A. Sali, collaborator). $550,000.
08/01/03 – 07/31/05	Towards a comprehensive map of protein-ligand interactions. California Institute for Quantitative Biomedical Research. $120,000.
08/03/03	IBM SUR award. Large-scale protein structure modeling and ligand docking (A. Sali, PI). $300,000.
09/15/03	Intel Inc. Award. Large-scale protein structure modeling and ligand docking (A. Sali, PI), $70,000.
11/01/03 – 10/31/06	Subnanometer structure based fold determination of biological complexes. NSF EIA-0324645 (W. Chiu, Bajaj, A. Sali, PIs), $520,000.
02/15/04 – 02/14/06	Hierarchical framework for structural biology. Award from the Sandler Program in Basics Sciences. $150,000.
04/01/04 – 03/31/09	Targeting Cysteine Proteases-Antiparasitic Chemotherapy. NIH P01 A135707 (J. McKerrow, PI; A. Sali, Co-PI). $140,000.
07/01/04 – 06/31/09	Deciphering Eznyme Specificity. NIH P01 GM71790 (J. Gerlt, PI; A. Sali, Co-PI). $400,000.
09/01/04 – 08/31/05	Structure-based discovery of Hsp90 inhibitors. UC Discovery Grant bio03-10401 (D. Agard, PI, A. Sali, PI). $63,960.
09/01/04– 08/31/08	Cataloguing the Human Salivary Proteome. NIH U01 DE016274 (S. Fisher, PI; A. Sali, Co-PI). $200.000.
07/01/05 – 06/30/10	Specialized Center for the Protein Structure Initiative. NIH U54 GM074929-01 (B. Stroud, PI; A. Sali, Co-PI). $360,000.
09/01/05 – 08/31/10	The Nuclear Information Pathway. NIH U54 RR022220 (M. Rout, PI; A. Sali, Co-PI). $1,259,214.
09/01/05 – 08/31/10	NYSGXRC: A Large Scale Center for the Protein Structure Initiative. NIH/NIGMS U54 GM62529 (S.K. Burley, PI; A. Sali, Co-PI). $1,600.000.
09/30/05 – 09/29/10	Center for Protein Folding Machinery. NIH PN2 EY016525-02 (W. Chiu, PI; A. Sali Co-PI). $200,000.

FELLOWSHIPS TO LAB MEMBERS

Howard Hughes Predoctoral Fellowship (R. Sanchez)
Howard Hughes Predoctoral Fellowship (F. Davis)

Burroughs Wellcome Fund Predoctoral Fellowship (R. Chiang)
Alfred P. Sloan Postdoctoral Fellowship (A. Stuart)
Burroughs Wellcome Fund Postdoctoral Fellowship (A. Fiser, M.A. Marti-Renom)
Rockefeller University Presidential Fellowship (M.A. Marti-Renom)
Charles Revson Foundation Postdoctoral Fellowship (A. Fiser, M.S. Madhusudhan)
NIH Postdoctoral Fellowship (R. Karchin)
Burroughs Wellcome Predoctoral Fellowship (M. Kim, R. Chiang)
DOE Predoctoral Fellowship (M. Peterson)
Genentech Award (M. Peterson)
Human Frontier Sciences Program Postdoctoral Fellowship (F. Foerster)

PUBLICATIONS

1. M. Renko, A. Sali, V. Turk, M. Pokorny, I. Kregar. "A neutral metalloproteinase from *Streptomyces rimosus*" *Vestnik Slovenskega Kemijskega Drustva* **32/2**, 161-173, 1985.

2. B. Lenarcic, A. Ritonja, A. Sali, M. Kotnik, V. Turk, W. Machleidt. "Properties and structure of human spleen stefin B - a low molecular weight protein inhibitor of cysteine proteinases" In: *Cysteine proteinases and their inhibitors*. Ed: V. Turk. Walter de Gruyter, Berlin, 473-487, 1986.

3. V. Turk, J. Brzin, B. Lenarcic, A. Sali, W. Machleidt. "Human stefins and cystatins: Their properties and structural relationships" In: *Cysteine proteinases and their inhibitors*. Ed: V. Turk. Walter de Gruyter, Berlin, 429-441, 1986.

4. M. Kotnik, A. Sali, J. Kos, B. Turk, V. Turk. "Nova metoda za hitro dolocanje kineticnih konstant pri interakciji encima s kompetitivnim inhibitorjem (A new method for rapid determination of kinetic constants for competitive inhibition of enzymes)" *Vestnik Slovenskega Kemijskega Drustva* **34**, 369-377, 1987.

5. A. Sali, V. Turk. "Prediction of the secondary structure of stefins and cystatins, the low Mr protein inhibitors of cysteine proteinases" *Biological Chemistry Hoppe-Seyler* **368**, 493-499, 1987.

6. T. Lah, I. Kregar, A. Sali, B. Lenarcic, M. Kotnik, V. Kostka, V. Turk. "Circular dichroism studies of different aspartyl proteinases and their interactions with pepstatin" *Periodicum Biologorum* **90**, 31-38, 1988.

7. V. Turk, R. Jerala, B. Lenarcic, A. Sali. "Structural and functional aspects of human cathepsins B" In: *Intracelullar Proteolysis: Mechanisms and Regulations*. Eds: N. Katunuma and E. Kominami. Japan Scientific Societies Press, Tokyo , 27-37, 1989.

8. A. Sali, B. Veerapandian, J.B. Cooper, S.I. Foundling, D.J. Hoover, T.L. Blundell. "High-resolution X-ray diffraction study of the complex between endothiapepsin and an oligopeptide inhibitor: The analysis of the inhibitor binding and description of the rigid body shift in the enzyme" *EMBO Journal* **8**, 2179-2188, 1989.

9. T.L. Blundell, G. Elliot, S.P. Gardner, T. Hubbard, S. Islam, M. Johnson, D. Mantafounis, P. Murray-Rust, J. Overington, J.E. Pitts, A.Sali, B.L. Sibanda, J. Singh, M.J.E. Sternberg, M.J. Sutcliffe, J.M. Thornton, P. Travers. "Protein engineering and design" *Philosophical Transactions of the Royal Society Lond. B* **324**, 447-460, 1989.

10. T.L. Blundell, D. Carney, T. Hubbard, M.S. Johnson, A. McLeod, J.P. Overington, A. Sali, M.S. Sutcliffe, P. Thomas. "Knowledge-based protein modelling and design" In: *Advances in Protein Design: International Workshop 1988 GBF Monographs 12*. Eds: H. Bloecker and J. Collins and R. D. Schmid and D. Schomburg. VCH Braunschweig, 39-43, 1989.

11. A. Sali, T.L. Blundell. "Definition of general topological equivalence in protein structures: A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming" *Journal of Molecular Biology* **212**, 403-428, 1990.

12. A. Sali, J.P. Overington, M.S. Johnson, T.L. Blundell. "From comparisons of protein sequences and structures to protein modelling and design" *Trends in Biochemical Sciences* **15**, 235-240, 1990.

13. B. Veerapandian, J.B. Cooper, A. Sali, T.L. Blundell. "X-ray analyses of aspartic proteinases III. Three-dimensional structure of endothiapepsin complexed with a transition-state isostere inhibitor of renin at 1.6Å resolution" *Journal of Molecular Biology* **216**, 1017-1029, 1990.

14. J. Overington, M.S. Johnson, A. Sali, T.L. Blundell. "Tertiary structural constraints on protein evolutionary diversity; Templates, key residues and structure prediction" *Proceedings of the Royal Society Lond. B* **241**, 132-145, 1990.

15. J.P. Overington, M.S. Johnson, C. Topham, A. McLeod, A. Sali, Z.-Y. Zhu, L. Sibanda, T.L. Blundell. "Applications of environment specific amino acid substitution tables to identification of key residues in protein tertiary structure" *Current Science* **59**, 867-874, 1990.

16. M.S. Johnson, A. Sali, T.L. Blundell. "Phylogenetic relationships from three-dimensional protein structures" *Methods in Enzymology* **183**, 670-690, 1990.

17. M.S. Johnson, J.P. Overington, A. Sali. "Knowledge-based protein modelling: Human plasma kallikrein and human neutrophil defensin" In: *Chemistry: Techniques Structure and Function*. Ed: J. J. Villafranca. Academic Press, Inc. San Diego, USA, 567-574, 1990.

18. M.S. Johnson, J. Overington, A. Sali, Z. Zhu, D. Donnelly, P. Thomas, A. McLeod, R. Goold, C. Topham, T.L. Blundell. "From comparative structure analysis to protein engineering: Knowledge-based protein modelling and design" *Fresnius Journal of Analytical Chem*istry **337**, 1-3, 1990.

19. T.L. Blundell, M.S. Johnson, J.P. Overington, A. Sali. "Knowledge-based protein modeling and the design of novel molecules" In: *Protein design and the development of new therapeutics and vaccines*. Eds: J. B. Hook and G. Poste. Plenum Press New York, 209-227, 1990.

20. T.L. Blundell, J.B. Cooper, A. Sali, Z.-Y. Zhu. "Comparisons of the sequences, 3-D structures and mechanism of pepsin-like and retroviral aspartic proteinases" In: *Structure and Function of the Aspartic Proteinases: Genetics, Structures, and Mechanisms*. Plenum Publishing Company New York, 443-453, 1991.

21. T.L. Blundell, J.B. Cooper, D. Donnelly, H. Driessen, Y. Edwards, F. Eisenmenger, C. Frazao, M. Johnson, K. Niefind, M.Newman, J. Overington, A. Sali, C. Slingsby, V. Nalini, Z.-Y. Zhu. "Patterns of sequence variation in families of homologous proteins" In: *Methods in Protein Sequence Analysis*. Eds: Jörnval, Höög and Gustavsson. Birkhauser Verlag, Basel , 373-385, 1991.

22. A. Sali, J.P. Overington, M.S. Johnson, T.L. Blundell. "From modelling homologous proteins to prediction of structure" In: *Protein design and the development of new therapeutics and vaccines*. Eds: J. M. Goodfellow and D. S. Moss. Ellis Horwood Ltd. New York , 231-245, 1991.

23. A. Sali, B. Veerapandian, J.B. Cooper, D.S. Moss, T. Hofmann, T.L. Blundell. "Domain flexibility in aspartic proteinases" *Proteins* **12**,158-170, 1992.

24. B. Veerapandian, J.B. Cooper, A. Sali, T.L. Blundell, R.L. Rosati, B.W. Dominy, D.B. Damon, D.J. Hoover. "Direct observation by X-ray analysis of a tetrahedral intermediate of aspartic proteinases" *Protein Science* **1**, 322-328, 1992.

25. Z.-Y. Zhu, A. Sali, T.L. Blundell. "A variable gap penalty function and feature weights for protein 3-D structure comparisons" *Protein Engineering* **5**, 43-51, 1992.

26. J. Overington, D. Donnelly, M.S. Johnson, A. Sali, T.L. Blundell. "Environment-specific amino acid substitution tables: Tertiary templates and prediction of protein folds" *Protein Science* **1**, 216-226, 1992.

27. M.S. Johnson, J.P. Overington, A. Sali, T.L. Blundell. "From the comparative analysis of proteins to similarity-based modelling" In: *Computer Modelling of Biomolecular Processes*. Eds: V. A. Ratner and N. A. Kolchanov. Nova Science Publishers, 6080 Jericho Turnpike, 191-196,

1992.

28. J.P. Overington, Z.-Y. Zhu, A. Sali, M.S. Johnson, R. Sowdhamini, G.V. Louie, T.L. Blundell. "Molecular recognition in protein families: A database of aligned three-dimensional structures of related proteins" *Biochemical Society Transactions* **21**, 597-604, 1993.

29. A. Sali, T.L. Blundell. "Comparative protein modelling by satisfaction of spatial restraints" *Journal of Molecular Biology* **234**, 779-815, 1993.

30. A. Sali, R. Matsumoto, H.P. McNeil, M. Karplus, R.L. Stevens. "Three-dimensional models of four mouse mast cell chymases. Identification of proteoglycan-binding regions and protease-specific antigenic epitopes" *Journal of Biological Chemistry* **268**, 9023-9034, 1993.

31. A. Sali, T.L. Blundell. "Comparative protein modelling by satisfaction of spatial restraints" In: Protein Structure by Distance Analysis. In: *Protein Structure by Distance Analysis*. Eds: H. Bohr and S. Brunak. IOS Press, Amsterdam, 64-86, 1994.

32. A. Sali, J.P Overington. "Derivation of rules for comparative protein modeling from a database of protein structure alignments" *Protein Science* **3**, 1582-1596, 1994.

33. A. Sali, E.I. Shakhnovich, M. Karplus. "Kinetics of protein folding: A lattice model study of the requirements for folding to the native state" *Journal of Molecular Biology* **235**, 1614-1636, 1994.

34. A. Sali, E.I. Shakhnovich, M. Karplus. "How does a protein fold?" *Nature* **369**, 248-251, 1994.

35. A. Dinner, A. Sali, M. Karplus, E. Shakhnovich. "Phase diagram of a model protein derived by exhaustive enumeration of the conformations" *Journal of Chemical Physics* **101**, 1444-1451, 1994.

36. M. Karplus, A. Sali. "Theoretical studies of protein folding and unfolding" *Current Opinion in Structural Biology* **5**, 58-73, 1995.

37. M. Karplus, A. Caflisch, A. Sali, E. Shakhnovich. "Protein dynamics: From the native to the unfolded state and back again" In: *Modelling of Biomolecular Structures and Mechanisms*. Eds: A. Pullman et al. Kluwer Academic Publishers, Dordrecht, Netherlands, 69-84, 1995.

38. A. Sali. "MODELER: Implementing 3D protein modeling" mc² Molecular Simulations Inc. Burlington, MA, 2, 5, 1995.

39. X. Wu, B. Knudsen, S.M. Feller, J. Zheng, A. Sali, D. Cowburn, H. Hanafusa, J. Kuriyan. "Structural basis for the specific interaction of lysine-containing proline-rich peptides with the

amino-terminal SH3 domain of c-Crk" *Structure* **15**, 215-226, 1995.

40. R. Matsumoto, A. Sali, N. Ghildyal, M. Karplus, R.L. Stevens. "Packaging of proteases and proteoglycans in the granules of mast cells and other hematopoietic cells. A cluster of histidines in mouse mast cell protease-7 regulates its binding to heparin serglycin proteoglycan" *Journal of Biological Chemistry* **270**, 19524-19531, 1995.

41. A. Sali. "Comparative protein modeling by satisfaction of spatial restraints" *Molecular Medicine Today* **1**, 270-277, 1995.

42. A. Sali, E. Shakhnovich, M. Karplus. "Protein Folding Studied by Monte Carlo Simulations" In: *Protein Folds: A Distance Based Approach*. Eds: H. Bohr and S. Brunak. CRC Press Inc., Florida, USA, 202-216, 1995.

43. A. Sali, E. Shakhnovich, M. Karplus. "Thermodynamics and kinetics of protein folding from lattice Monte Carlo simulations" In: *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. Eds: D. Shalloway, G. Xue and P. Pardalos. American Mathematical Society, **23**, 199-213, 1995.

44. A. Sali, L. Potterton, F. Yuan, H. van Vlijmen, M. Karplus. "Evaluation of comparative protein structure modeling by MODELLER" *Proteins* **23**, 318-326, 1995.

45. A. Sali. "Modelling mutations and homologous proteins" *Current Opinion in Biotechnology* **6**, 437-451, 1995.

46. Y. Sheng, A. Sali, H. Herzog, J. Lahnstein, S.A. Krilis. "Site-directed mutagenesis of recombinant human β_2-glycoprotein Iidentifies a cluster of lysine residues that are critical for phospholipid binding and anti-cardiolipin antibody activity" *Journal of Immunology* **157**, 3744-3751, 1996.

47. N. Ghildyal, D.S. Friend, R.L. Stevens, K.F. Austen, C. Huang, J. Penrose, A. Sali, M.F. Gurish. "Fate of two mast cell tryptases following passive systemic anaphylaxis of BALB/c and V3 mastocytosis mice. Prolonged retention of exocytosed mMCP-6 in connective tissues and rapid accumulation of enzymatically active mMCP-7 in the blood chymases." *Journal of Experimental Medicine* **184**, 1061-1073, 1996.

48. L.Z. Xu, R. Sánchez, A. Sali, N. Heintz. "Ligand specificity of brain lipid binding protein" *Journal of Biological Chemistry* **271**, 24711-24719, 1996.

49. A.R. Dinner, A. Sali, M. Karplus. "The folding mechanism of larger model proteins: Role of native structure" *Proceedings of the National Academy of Sciences USA* **98**, 8356-8361, 1996.

50. S. Wu, H. deLencastre, A. Sali, A. Tomasz. "A phosphoglucomutase-like gene essential for the optimal expression of methicillin resistance in *Staphylococcus aureus*: Molecular cloning and DNA sequencing" *Microbial Drug Resistance* **2**, 277-286, 1996.

51. M. Russel, N.A. Linderoth, A. Sali. "Filamentous phage assembly: Variation on a protein export theme" *Gene* **192**, 23-32, 1997.

52. C. Huang, G.W. Wong, N. Ghildyal, M.F. Gurish, A. Sali, R. Matsumoto, W.-T. Qiu, R.L. Stevens. "The tryptase, mouse mast cell protease 7, exhibits anticoagulant activity in vivo and in vitro due to its ability to degrade fibrinogen in the presence of the diverse array of protease inhibitors in plasma" *Journal of Biological Chemistry* **272**, 31885-31893, 1997.

53. R. Sánchez, A. Sali. "Comparative protein modeling as an optimization problem" *Journal of Molecular Structure (Theochem)* **398**,489-496, 1997.

54. R. Sánchez, A. Sali. "Advances in comparative protein-structure modeling" *Current Opinion in Structural Biology* **7**, 206-214, 1997.

55. D. Koulich, M. Orlova, A. Malhotra, A. Sali, S.A. Darst, S. Borukhov. "Domain organization of *Escherichia coli* transcript cleavage factors GreA and GreB" *Journal of Biological Chemistry* **272**, 7201-7210, 1997.

56. J.E. Hunt, D.S. Friend, M.F. Gurish, E. Feyfant, A. Sali, C. Huang, N. Ghildyal, S. Stechshulte, F.K. Austen, R.L. Stevens. "Mouse mast cell protease (mMCP) 9, a novel member of the chromosome 14 family of serine proteases that is selectively expressed in uterine mast cells" *Journal of Biological Chemistry* **46**, 29158-29166, 1997.

57. R. Sánchez, A. Sali. "Evaluation of comparative protein structure modeling by MODELLER-3" *Proteins* **Suppl. 1**, 50-58, 1997.

58. B. Guenther, R. Onrust, A. Sali, M. ODonnell, J. Kuriyan. "Crystal structure of the ' subunit of the clamp-loader complex of *E. coli* DNA polymerase III" *Cell* **91**, 335-345, 1997.

59. R. Sánchez, A.Ya. Badretdinov, E. Feyfant, A. Sali. "Homology protein structure modeling" *Transactions of the American Crystallographic Association* **32**, 81-91, 1997.

60. C.M. Dobson, A. Sali, M. Karplus. "Protein folding: A perspective from theory and experiment" *Angewandte Chemie International Edition* **37**,868-893, 1998.

61. C. Huang, A. Sali, R.L. Stevens. "Regulation and function of mast cell proteases in inflammation" *Journal of Clinical Immunology* **18**, 169-183, 1998.

62. D.A. Kandiah, A. Sali, Y. Sheng, E.J. Victoria, D.M. Marquis, S.M. Coutts, S.A. Krilis.

"Current insights into the antiphospholipid syndrome: Clinical, immunological and molecular aspects" *Advances in Immunology* **70**, 507-563, 1998.

63. A. Gutin, A. Sali, V. Abkevich, M. Karplus, E.I. Shakhnovich. "Temperature dependence of the folding rate in a simple protein model: Search for a `glass transition'" *Journal of Chemical Physics* **108**, 6466-6483, 1998.

64. E. Wolf, A. Vassilev, Y. Makino, A. Sali, Y. Nakatani, S.K. Burley. "Crystal structure of a GCN5-related N-acetyltransferase: *Serratia marcescens* aminoglycoside 3-N-acetyltransferase" *Cell* **94**, 51-61, 1998.

65. R. Sánchez, A. Sali. "Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome" *Proceedings of the National Academy of Sciences USA* **95**,13597-13602, 1998.

66. A. Sali. "100,000 protein structures for the biologist" *Nature Structural Biology* **5**, 1029-1032, 1998.

67. R. Sánchez, A. Sali. "MODBASE: A database of comparative protein structure models" *Bioinformatics* **15**, 1060-1061, 1999.

68. R. Sánchez, A. Sali. "Comparative protein structure modeling in genomics" *Journal of Computational Physics* **151**, 388-401, 1999.

69. G. Wu, A. Fiser, B. terKuile, A. Sali, M. Muller. "Convergent evolution of *Trichomonas vaginalis* lactate dehydrogenase from malate dehydrogenase" *Proceedings of the National Academy of Sciences USA* **96**, 6285-6290, 1999.

70. T. Nagata, V. Gupta, W-Y. Kim, A. Sali, B.T. Chait, K. Shigesada, Y. Ito, M.H. Werner. "Immunoglobulin motif DNA recognition and heterodimerization for the PEBP2/CBF Runt--domain" *Nature Structural Biology* **6**, 615-619, 1999.

71. J.M. Miwa, I. Ibanez-Tallon, G.W. Crabtree, R. Sánchez, A. Sali, L.W. Role, N. Heintz. "lynx1, an endogenous toxin-like modulator of nicotinic acetylcholine receptors in the mammalian CNS" *Neuron* **23**, 105-114, 1999.

72. G.W. Wong, Y. Tang, E. Feyfant, A. Sali, C. Huang, R.L. Stevens. "Identification of a new tryptase in mice and humans which possesses a novel C-terminal, membrane spanning segment" *Journal of Biological Chemistry* **274**, 30784-30793, 1999.

73. S.K. Burley, S.C. Almo, J.B. Bonanno, M. Capel, M.R. Chance, T. Gaasterland, D. Lin, A. Sali, F.W. Studier, S. Swaminathan. "Structural genomics: beyond the Human Genome Project" *Nature Genetics* **23**, 151-157, 1999.

74. A. Sali, J. Kuriyan. "Challenges at the frontiers of structural biology" *Trends in Biochemical Sciences* **22**, M20-M24, 1999.

75. A. Sali. "Functional links between proteins" *Nature* **402**, 23-26, 1999.

76. C. Huang, G. Morales, A. Vagi, K. Chanasyk, M. Ferrazzi, C. Burklow, W.T. Qiu, E. Feyfant, A. Sali, R.L. Stevens. "Formation of enzymatically active, homotypic and heterotypic tetramers of mouse mast cell tryptases" *Journal of Biological Chemistry* **275**, 351-358, 2000.

77. M.A. Martí-Renom, A. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali. "Comparative protein structure modeling of genes and genomes" *Annual Review of Biophysics and Biomolecular Structure* **29**, 291-325, 2000.

78. R. Sánchez, U. Pieper, N. Mirkovic, P.I.W. deBakker, E. Wittenstein, A. Sali. "MODBASE, a database of annotated comparative protein structure models" *Nucleic Acids Research* **28**, 250-253, 2000.

79. A.R. Dinner, A. Sali, L.J. Smith, C.M. Dobson, M. Karplus. "Understanding protein folding via free-energy surfaces from theory and experiment" *Trends in Biochemical Sciences* **25**, 331-339, 2000.

80. R. Sánchez, A. Sali. "Comparative protein structure modeling: Introduction and practical examples with MODELLER" In: *Protein Structure Prediction: Methods and Protocols*. Ed: D. M. Webster. Humana Press, 97-129, 2000.

81. G. Wu, H.G. Morrison, A. Fiser, A.G. McArthur, A. Sali, M.L. Sogin, M. Muller. "Core histones of the amitochondriate protist *Giardia lamblia*" *Molecular Biology and Evolution* **17**, 1156-1163, 2000.

82. A. Fiser, R.K. Do, A. Sali. "Modeling of loops in protein structures" *Protein Science* **9**, 1753-1773, 2000.

83. W. Liedtke, Y. Choe, M.A. Martí-Renom, A.M. Bell, C.S. Denis, A. Sali, A.J. Hudspeth, H.M. Friedman, S. Heller. "Vanilloid receptor-related mechanically-activated channel (VR-MARC), a candidate vertebrate osmoreceptor" *Cell* **103**, 525-535, 2000.

84. S. Jin, S. Martinek, W.S. Joo, J.R. Wortman, N. Mirkovic, A. Sali, M.D. Yandell, N.P. Pavletich, M.W. Young, A.J. Levine. "Identification and characterization of a p53 homologue in *Drosophila melanogaster*" *Proceedings of the National Academy of Sciences USA* **97**, 7301-7306, 2000.

85. R. Sánchez, U. Pieper, F. Melo, N. Eswar, M.A. Martí-Renom, M.S. Madhusudhan, N. Mirkovic, A. Sali. "Protein structure modeling for structural genomics" *Nature Structural Biology*

7, 986-990, 2000.

86. A. Fiser, R. Sánchez, F. Melo, A. Sali. "Comparative protein structure modeling" In: *Computational Biochemistry and Biophysics*. Eds: M. Watanabe, B. Roux, A. MacKerell and O. Becker. Marcel Dekker, 275-312, 2000.

87. J. Vernal, A. Fiser, A. Sali, M. Müller, J.J. Cazzulo, C. Nowicki. "Probing the specificity of a trypanosomal aromatic fa-hydroxy acid dehydrogenase by site-directed mutagenesis" Biochemical and Biophysical Research Communications **293**, 633-639, 2002.

88. C.M. Groft, R. Beckmann, A. Sali, S.K. Burley. "Crystal structures of ribosome anti-association factor IF6" *Nature Structural Biology* **7**, 1156-1164, 2000.

89. S. Gopal, M. Schroeder, U. Pieper, A. Sczyrba, G. Aytekin-Kurban, S. Bekiranov, E.J. Fajardo, N. Eswar, R. Sánchez, A. Sali, T.Gaasterland. "Homology-based annotation yields 1,042 new candidate genes in the *Drosophila melanogaster* genome" *Nature Genetics* **27**, 337-340, 2001.

90. F. Melo, R. Sánchez, A. Sali. "Statistical potentials for fold assessment" *Protein Science* **11**, 430-448, 2002.

91. V.A. Eyrich, M.A. Martí-Renom, D. Przybylski, M.S. Madhusudhan, A. Fiser, F. Pazos, A. Valencia, A. Sali, B. Rost. "EVA: continuous automatic evaluation of structure prediction servers" *Bioinformatics* **17**, 1242-1243, 2001.

92. A.C. Stuart, V.A. Illyin, A. Sali. "LigBase: A database of families of aligned ligand binding sites in known protein sequences and structures" *Bioinformatics* **18**, 200-201, 2002.

93. M.A. Martí-Renom, V.A. Illyin, A. Sali. "DBAli: A database of alignments and tools for benchmarking protein sequence-structure comparison" *Bioinformatics* **17**, 746-747, 2001.

94. M.A. Martí-Renom, M.S. Madhusudhan, A. Fiser, B. Rost, A. Sali. "Reliability of assessment of protein structure prediction methods" *Structure* **10**, 435-440, 2002.

95. G.W. Wong, L. Li, M.S. Madhusudhan, S.A. Krilis, M.F. Gurish, M.E. Rothenberg, A. Sali, R.L. Stevens. "Tryptase 4: A new member of the chromosome 17 family of mouse serine proteases" *Journal of Biological Chemistry* **276**, 20648-20658, 2001.

96. L.G. Barrientos, R. Campos-Olivas, J.M. Louis, A. Fiser, A. Sali, A.M. Gronenborn. "H-1, C-13, N-15 resonance assignments and fold verification of a circular permuted variant of the potent HIV-inactivating protein cyanovirin-N" *Journal of Biomolecular NMR* **19**, 289-290, 2001.

97. Y. Sheng, S.A. Krilis, A. Sali. "Site-directed mutagenesis of recombinant human

glycoprotein I. Effect of phospholipid bindingand anticardiolipin antibody activity" *Annals of the New York Academy of Sciences* **815**, 331-333, 1997.

98. A. Sali, M.A. Marti-Renom, M.S. Madhusudhan, A. Fiser, B. Rost. "Reply to Moult et al." *Structure* **10**, 292-293, 2002.

99. A. Sali. "Target practice" *Nature Structural Biology* **8**, 482-484, 2001.

100. Min-yi Shen, F.P. Davis, A. Sali. The optimal size of a globular protein domain: a simple sphere-packing model. *Chemical Physics Letters* in press **405**, 224-228, 2005.

101. C.M.T. Spahn, R. Beckmann, N. Eswar, P.A. Penczek, A. Sali, G. Blobel, J. Frank. "Structure of the 80S ribosome from *Saccharomyces cerevisiae* - tRNA-ribosome and subunit-subunit interactions" *Cell* **107**, 373-386, 2001.

102. R. Beckmann, C.M.T. Spahn, N. Eswar, J. Helmers, P.A. Penczek, A. Sali, J. Frank, G. Blobel. "Architecture of the protein-conducting channel associated with the translating 80S ribosome" *Cell* **107**, 361-372, 2001.

103. G.W. Wong, S. Yasuda, M.S. Madhusudhan, L. Li, Y. Yang, S.A. Krilis, A. Sali, R.L. Stevens. "Human tryptase epsilon (PRSS22), a new member of the chromosome 16p13.3 family of human serine proteases expressed in airway epithelial cells" *Journal of Biological Chemistry* **276**, 49169-49182, 2001.

104. A. Fiser, M. Feig, C.L. Brooks III, A. Sali. "Evolution and physics in comparative protein structure modeling" *Accounts of Chemical Research*, **35**, 413-421, 2002.

105. A. Fiser, A. Sali. "MODELLER: generation and refinement of homology-based protein structure models" In: *Methods in Enzymology Vol. 374*. Eds: C.W. Carter, R.M. Sweet. Academic Press, San Diego, Chapter 20, 463-493, 2003.

106. D. Baker, A. Sali. "Protein structure prediction and structural genomics" *Science* **294**, 93-96, 2001.

107. V.A. Ilyin, U. Pieper, A.C. Stuart, M.A. Martí-Renom, L. McMahan, A. Sali. "ModView, visualization of multiple protein sequences and structures" *Bioinformatics* **19**, 165-166, 2002.

108. U. Pieper, N. Eswar, A.C. Stuart, V.A. Ilyin, A. Sali. "MODBASE, a database of annotated comparative protein structure models" *Nucleic Acids Research* **30**, 255-259, 2002.

109. M. Borodovsky, E. Koonin, C. Burge, J. Fickett, J. Logsdon, A. Sali, G. Stormo, I. Zhulin. "The third Georgia Tech – Emory international conference on bioinformatics: in silico biology; bioinformatics after human genome (November 15-18, 2001, Atlanta, Georgia,USA)"

Bioinformatics **17**, 859-861, 2001.

110. P. Baldi, M. Borodovsky, S. Brunak, C. Burge, J. Fickett, S. Henikoff, E. Koonin, A. Sali, C. Sander, G. Stormo. "The Second Georgia Tech International Conference on Bioinformatics: Sequence, Structure and Function" *Bioinformatics* **15**, 865-866, 1999.

111. M.R. Chance, A.R. Bresnick, S.K. Burley, J.S. Jiang, C.D. Lima, A. Sali, S.C. Almo, J.B. Bonanno, J.A. Buglino, S. Boulton, H. Chen, N. Eswar, G. He, R. Huang, V. Ilyin, L. McMahan, U. Pieper, S. Ray, M. Vidal, L.K. Wang. "Structural Genomics: A pipeline for providing structures for the biologist" *Protein Science* **11**, 723-738, 2002.

112. K.R. Rajashankar, M.R. Chance, S.K. Burley, J. Jiang, S.C. Almo, A. Bresnick, T. Dodatko, R. Huang, G. He, H. Chen, M.Sullivan, J. Toomey, R.A. Thirumuruhan, W.A. Franklin, A. Sali, U. Pieper, N. Eswar, V. Ilyin, L. McMahan. "Activity Report of NSLS 2001" **2**, 28-32, 2002.

113. C.M. Groft, R. Beckmann, A. Sali, S.K. Burley. "Response to Paoli" *Nature Structural Biology* **8**, 745, 2001.

114. Y. Yang, L. Li, G.W. Wong, S.A. Krilis, M.S. Madhusudhan, A. Sali, R.L. Stevens. "RasGRP4: a new mast cell-restricted, ras guanine nucleotide releasing protein with calcium- and diacylglycerol-binding motifs" *Journal of Biological Chemistry*, **277**, 25756-25774,2002.

115. M.A. Marti-Renom, B. Yerkovich, A. Sali. "Comparative protein structure prediction", *Current Protocols in Protein Science* **1**, 2.9.1-2.9.22, 2002.

116. G.M. Iverson, S. Reddel, E.J. Victoria, K.A. Cockerill, WangY-X, M.A. Marti-Renom, A. Sali, D.M. Marquis, S.A. Krillis, M.D.Linnik. "Use of single point mutations in domain I of β2-Glycoprotein I to determine fine antigenic specificity of antiphospholipid autoantibodies" *Journal of Immunology* **169**, 7097-7103, 2002.

117. M.A. Marti-Renom, B. Yerkovich, A. Sali. "Modeling protein structure from its sequence" *Current Protocols in Bioinformatics John Wiley & Sons, Inc.* **V. 5 January Issue**, Pages 5.1.1-5.1.32, 2003.

118. M. Karplus, A. Sali, E. Shakhnovich. "Kinetics of protein folding" Nature **373**, 664-665, 1995.

119. A. Fiser, A. Sali. "Comparative Protein Structure Modeling" In: *Protein Structure*. Ed: Daniel Chasman. *Marcel Dekker, Inc. New York*, 167-206, 2003.

120. A. Sali, R. Glaeser, T. Earnest, W. Baumeister. "From words to literature in structural proteomics" *Nature* **422**, 216-225, 2003.

121. I.Y.Y. Koh, V.A. Eyrich, M.A. Marti-Renom, D. Przybylski, M.S. Madhusudhan, E. Narayanan, O. Grana, F. Pazos, A. Valencia, A. Sali, B. Rost. "EVA: evaluation of protein structure prediction servers" *Nucleic Acids Research* **31**, 3311-3315, 2003.

123. B. John, A. Sali. "Comparative Protein Structure Modeling by Iterative Alignment, Model Building, and Model Assessment" *Nucleic Acids Research* **31**, 1982-1992, 2003.

124. A. Rossi, Q. Deveraux, B. Turk, A. Sali. "Comprehensive search for cysteine cathepsins in the human genome" *Biological Chemistry* **385**, 363-372, 2004.

125. N. Eswar, B. John, N. Mirkovic, A. Fiser, V.A. Ilyin, U. Pieper, A.C. Stuart, M.A. Marti-Renom, M.S. Madhusudhan, B. Yerkovich, A. Sali. "Tools for comparative protein structure modeling and analysis" *Nucleic Acids Research* **31**, 3375-3380, 2003.

126. H. Gao, J. Sengupta, M. Valle, A. Korostelev, N. Eswar, S.M. Stagg, P. VanRoey, R.K. Angrawal, S.C. Harvey, A. Sali, M.S. Chapman, J. Frank. "Study of the Structural Dynamics of the *E. coli* 70S Ribosome Using Real Space Refinement" *Cell* **113**, 789-801, 2003.

127. M.A. Marti-Renom, M.S. Madhusudhan, A. Sali. "Alignment of Protein Sequences by Their Profiles" *Protein Science* **13**, 1071-1087, 2004.

128. S. Lee, E. Shen, A. Fiser, A. Sali, S. Guo. "The zebrafish forkhead transcription factor Foxi1 specifies epibranchial placode-derived sensory neurons", *Development* **130,** 2669-2679, 2003.

129. M.A. Marti-Renom, A. Fiser, M.S. Madhusudhan, B. John, A. Stuart, N. Eswar, P. Pieper, M-Y. Shen, A. Sali. "Modeling protein structure from its sequence" *Current Protocols in Bioinformatics John Wiley & Sons, Inc.* **V. 5 November Issue**, Pages 5.1.1-5.1.32, 2003.

130. Andrej Sali. "NIH Workshop on Structural Proteomics of Biological Complexes" *Structure* **11**, 1043–1047, 2003.

131. F. Alber, N. Eswar, A. Sali. "Structure determination of macromolecular complexes by experiment and computation" in *Practical Bioinformatics*, Ed: J. Bujnicki. *Springer-Verlag, Berlin Heidelberg* **15**, 73-96, 2004.

132. B. John, A. Sali. "Detection of homologous proteins by an intermediate sequence search" *Protein Science* **13**, 54-62, 2004.

133. E. Feyfant, A. Fiser, A. Sali. "Modeling of mutations in protein structures", in preparation.

134. U. Pieper, N. Eswar, H. Braberg, M.S. Madhusudhan, F. Davis, A.C. Stuart, N. Mirkovic, A. Rossi, M.A. Marti-Renom, A. Fiser, B. Webb, D. Greenblatt, C. Huang, T. Ferrin, A. Sali. "MODBASE, a database of annotated comparative protein structure models, and associated

resources", *Nucl. Acids. Res.* **32**, D217-D222, 2004.

135. A. Fiser, A. Sali. "ModLoop: automated modeling of loops in protein structures" *Bioinformatics* **19**, 1-2, 2003.

136. S.M. Maurer, A. Rai, A. Sali. "Finding Cures for Tropical Diseases: Is Open Source An Answer?" Ed: Lynn Yarris, Lawrence Berkeley Laboratory. *Biotechnology: Essays From Its Heartland,* **June 2004**, 33-37, 2004

137. M.R. Chance, A. Fiser, A. Sali, U. Pieper, N. Eswar, G. Xu, E. Fajardo, T. Radhakannan, N. Marinkovic. "High-Throughput Computational and Experimental Techniques in Structural Genomics" *Genome Research,* **14**, 2145-2154, 2004.

138. M. Jacobson, A. Sali. "Comparative Protein Structure Modeling and Its Applications to Drug Discovery" Ed: John Overington. *Annual Reports in Medicinal Chemistry Inpharmatica Ltd., 60 Charlotte St, London, W1T2NU, UK*, **39**, 259-276, 2004.

139. R.B. Russell, F. Alber, P. Aloy, F.P. Davis, D. Korkin, M. Pichaud, M. Topf, A. Sali. "A structural perspective on protein-protein interactions and complexes" *Current Opinion in Structural Biology* **14**, 313-324, 2004.

140. D. Devos, S. Dokudovskaya, F. Alber, R. Williams, B.T. Chait, A. Sali, M.P. Rout. "Components of Coated Vesicles and Nuclear Pore Complexes Share a Common Molecular Architecture" *PLoS Biology,* December 2004, **Vol. 2 Issue 12**, 1-9, 2004.

141. J. Espadaler, R. Aragüés, N. Eswar, M.A. Marti-Renom, E. Querol, F.X. Avilés, A. Sali, B. Oliva. "Detecting Remotely Related Proteins by Their Interactions and Sequence Similarity", *PNAS* **102**, 7151-7156, 2005.

142. M. Topf, M.L. Baker, B. John, W. Chiu, A. Sali. "Structural Characterization of Components of Protein Assemblies by Comparative Modeling and Electron Cryo-Microscopy", *J. Struct. Biol.* **149**, 191-203, 2005.

143. R. Karchin, L. Kelly, A. Sali. "Improving functional annotation of non-synonymous SNPs with information theory", *Pacific Symposium on Biocomputing*, 397-408, 2005.

144. M.S. Madhusudhan, M.A. Marti-Renom, N. Eswar, B. John, U. Pieper, R. Karchin, S. Min-yi, A. Sali. "Comparative Protein Structure Modeling" In: *The Proteins Protocols Handbook. Ed: J.M. Walker. Humana Press Inc. Totowa, NJ*, 831-860, 2005.

145. S.M. Maurer, A. Rai, A. Sali. "Finding Cures for Tropical Diseases: Is Open Source and Answer?". *PloS Medicine* **1**, 180-183, 2004.

146. F. Alber, M. Kim, A. Sali. "Structural characterization of assemblies from overall shape and subcomplex compositions", *Structure* **13**, 435,-445, 2005.

147. F.P. Davis, A. Sali. "PIBASE: A Comprehensive Database of Structurally Defined Protein Domain Interfaces", *Bioinformatics* **21**, 1901-1907, 2005.

148. R. Karchin, M. Diekhans, L. Kelly, D.J. Thomas, U. Pieper, N. Eswar, D. Haussler, A. Sali. LS-SNP: "Large-scale annotation of coding non-synonymous SNPs based on multiple information sources", *Bioinformatics* **21**, 2814-2820, 2005.

149. S. Yasuda, N. Morokawa, G.W. Wong, A. Rossi, M.S. Madhusudhan, A. Sali, Y.S. Askew, G.A. Silverman, S.A. Krilis, R.L. Stevens. "Urokinase-type plasminogen activator is a preferred substrate of the human epithelium serine protease tryptase epsilon/PRSS22". *Blood* **105**, 3893-3901 , 2005.

150. J. Dvorak, M. Delcroix, A. Rossi, V. Vopalensky, M. Pospisek, M. Sedinova, L. Mikes, M. Sajid, A. Sali, J.H. Mckerrow, P. Horak, C.R. Caffrey. "Multiple cathepsin B isoforms in schistosomula of Trichobilharzia regenti: Identification, characterization and putative role in migration and nutrition". *International Journal of Parasitology* **35**, 895-910, 2005.

151. A. Sali, W. Chiu. "Marcromolecular Assemblies Highlighted". *Structure* **13**, 339-341, 2005.

152. D. Korkin, F.P. Davis, A. Sali. "Localization of protein binding sites within families of proteins" *Protein Science* **14**, 2350-2360, 2005.

153. D. Devos, S. Dokudovskaya, R. Williams, F. Alber, N. Eswar, B.T. Chait, M.P. Rout, A. Sali. "Simple fold composition and modular architecture of the nuclear pore complex" *PNAS* **14**, 2172-2177, 2006.

154. M.S. Madhusudhan, M.A. Marti-Renom, R. Sanchez, A. Sali. "Variable gap penalty for protein sequence-structure alignment" *Protein Engineering Design & Selection*, in press.

155. Min-yi Shen, F.P. Davis, A. Sali. "The optimal size of a globular protein domain: a simple sphere-packing model" *Chemical Physics Letters* **405,** 224-228, 2005.

156. J.B. Bonanno, S.C. Almo, A. Bresnick, M.R. Chance, A. Fiser, S. Swaminathan, J. Jiang, F.W. Studier, L. Shapiro, C. Lima, T. Gaasterland, A. Sali, K. Bain, I. Feil, X. Gao, D. Lorimer, A. Ramos, M. Sauder, S. Wasserman, S. Emtage, K. DAmico, S.K. Burley. "New York-Structural GenomiX Research Consortium (NYSGXRC): a large scale center for the protein structure initiative" *Journal of Structural and Functional Genomics* **6**, 225-232, 2005.

157. M. Topf, A. Sali. "Combining Electron Microscopy and Comparative Protein Structure Modeling" *Current Opinion in Structural Biology* **15**, 578-585, 2005.

158. S.A. McMahon, J.L. Miller, J.A. Lawton, D.E. Kerkow, A. Hodes, M.A. Marti-Renom, S. Doulatov, E. Narayanan, A. Sali, J.F. Miller, P. Ghosh. "The C-type lectin fold as an evolutionary solution for massive sequence variation" *Nature Structural & Molecular Biology* **12**, 886-892, 2005.

159. S.M. Maurer, A. Rai, A. Sali. "Finding Cures for Tropical Diseases: Is Open Source An Answer?" *Minnesota Journal of Law, Science & Technology* **6**, 169-175, 2004.

160. U. Pieper, N. Eswar, F.P. Davis, H. Braberg, M.S. Madhusudhan, A. Rossi, M. Marti-Renom, R. Karchin, B. Webb, D. Eramian, M.Y. Shen, L. Kelly, F. Melo, A. Sali. "MODBASE, a database of annotated comparative protein structure models and associated resources" *Nucleic Acids Research* **34**, D291-D295, 2006.

161. M. Topf, M.L. Baker, M.A. Marti-Renom, W. Chiu, A. Sali. "Refinement of Protein tructures by Iterative Comparative Modeling and CryoEM Density Fitting" *Journal of Molecular Biology* **357**, 1655-1668, 2006.

162. N. Eswar, A. Sali. "Comparative Modeling of Drug Target Proteins" In: *Comprehensive Medicinal Chemistry II.* Ed: J. Taylor and D. Triggle. *Elsevier,* in press.

163. D. Eramian, M.Y. Shen, D. Devos, F. Melo, A. Sali, M.A. Marti-Renom. "A composite score for predicting errors in protein structure models" *Protein Science* **15**,1653-66, 2006.

164. S. Dokudovskaya, R. Williams, D. Devos, A. Sali, B.T. Chait, M. Rout. "Protease Accessibility Laddering (PAL) - A Proteomic Tool for Probing Protein Structure" *Structure* **14**, 653-660, 2006.

165. N. Eswar, M.A. Marti-Renom, B. Webb, M.S. Madhusudhan, D. Eramian, M.Y. Shen, U. Pieper, A. Sali. "Comparative Protein Structure Modeling with MODELLER" *Current Protocols in Bioinformatics* John Wiley & Sons, Inc. Supplement 15, 5.6.1-5.6.30, 2006.

166. A. Rossi, M.A. Marti-Renom, A. Sali. "Localization of binding sites in protein structures by optimization of a composite scoring function". *Protein Science* **15(10)**, 2366-80, 2006.

167. F.P. Davis, H. Braberg, M.Y. Shen, U. Pieper, A. Sali, M.S. Madhusudhan. "Protein complex compositions predicted by structural similarity" *Nucleic Acids Research* 31;34(10), 2943-52, 2006.

168. V.D. Winn, R. Haimov-Kochman, A.C. Paquet, Y.J. Yang, M.S. Madhusudhan, M. Gormley, V. Feng, D.A. Bernlohr, S. McDonagh, L. Pereira, A. Sali, S.J. Fisher. "Gene Expression Profiling of the Human Maternal-Fetal Interface Over Gestation Reveals"

Endocrinology **148**, 1059-79, 2006.

169. T.D. Nguyen, J.M. Gow, L.W. Chinn, L. Kelly, H. Jeong, C.C. Huang, D. Stryke, M. Kawamoto, S.J. Johns, E. Carlson, T. Taylor, T.E. Ferrin, A. Sali, K.M. Giacomini, D.L. Kroetz. "PharmGKB Submission Update: IV. PMT Submissions of Genetic Variations in ATP-Binding Cassette Transporters to the PharmGKB Network" *Pharmacological Reviews* **58**, 1-2, 2006.

170. H.M. Berman, S.K. Burley, W. Chiu, A. Sali, A. Adzhubei, P.E. Bourne, S.H. Bryant, J. Roland L. Dunbrack, K. Fidelis, J. Frank, A. Godzik, K. Henrick, A. Joachimiak, B. Heymann, D. Jones, J.L. Markley, J. Moult, G.T. Montelione, C. Orengo, M.G. Rossmann, B. Rost, H. Saibil, T. Schwede, D.M. Standley, and J.D. Westbrook , "Outcome of a workshop on archiving structural models of biological macromolecules". *Structure,* **14**, 1211-1217, 2006.

171. A. Colubri, A. Jha, Min-yi Shen, A. Sali, R S. Berry, T. Sosnick, K.F. Freed. "Minimalist representations and the importance of nearest neighbor effects in protein folding simulations" *Journal of Molecular Biology* **363(4)**, 835-57, 2006.

172. M. Shen, A. Sali. "Statistical Potential for Assessment and Prediction of Protein Structure" *Protein Science,* 15, 2507 – 2524, 2006.

173. R. Karchin, A.N.A. Monteiro, S. V. Tavtigian, M. A. Carvalho, A. Sali "Functional impact of missense variants in BRCA1 predicted by supervised learning". *PLoS Computational Biology* **3(2): e26**, 2007.

174. L. Kelly, R. Karchin, A. Sali. "Protein interactions and disease phenotypes in the ABC transporter superfamily". *Pacific Symposium on Biocomputing.* Accepted.

175. D. Korkin, F.P. Davis, F. Alber, T. Luong, M. Shen, V. Lucic, M. B. Kennedy, A. Sali "Structural Modeling of Protein Interactions by Analogy: Application to PSD-95". *PLoS Computational Biology*, **2(11); e153**, 1365-1376, 2006.

176. M.A. Carvalho, S.M. Marsillac, R. Karchin, S. Manoukian, S. Grist , R. Swaby, T. P. Urmenyi, E. Rondinelli, R. Silva, L. Gayol, L. Baumbach, R. Sutphen, J. L. Pickard, K. L. Nathanson, A. Sali, D. Goldgar, F. J. Couch, P. Radice, A.N.A. Monteiro. "Determination of cancer risk associated with germline BRCA1 missense variants by functional analysis". *Cancer Research* **15;67(4)**, 1494-501, 2007.

177. S.X. Wang, K. C. Pandey, J. Scharfstein, J. Whisstock, R. K. Huang, J. Jacobelli, R. J. Fletterick, P. J. Rosenthal, M. Abrahamson, A. Rossi, A. Sali, J. H. McKerrow. "The Structure of Chagasin Reveals New Relationships in the Molecular Evolution of Protease Inhibitors". Sunmitted, 2006.

178. C.Lima, J.Puglisi, A.Sali, L. Szewczak. "Editorial". *Structure* 14. 2006

179. N. Soranzo, L. Kelly, L. Martinian, M-W. Burley, M. Thom, A. Sali, D. L. Kroetz, D. B. Goldstein, S. M. Sisodiya. "Lack of support for a role for RLIP76 (RALBP1) in response to treatment or predisposition to epilepsy". *Epilepsia*, in press, 2006.

180. K. Bajaj; M.S. Madhusudhan; P. Chakrabarti; C. Ramakrishnan; A. Sali; R. Varadarajan "Stereochemical criteria for prediction of the effects of Proline mutations on protein stability". Submitted, 2006.

181. N.Eswar, D.Eramian, B.Webb, M.Shen, A.Sali. "Protein Structure Modeling With MODELLER" in Methods in Molecular Biology, Humana Press, in press.

182. Karchin R, Parmigiani G, Couch F, Agarwal M, Sali A, Ziegler J & Beattie M.S. "Classification of BRCA2 Unclassified Variants using protein and family history information." Submitted.

183. R.Aragues, A.Sali, M.A.Marti-Renom, B.Oliva. "Gene Essentiality is Correlated with the Number of Inferred Protein Interacting Motifs", Submitted.

184. M.A.Marti-Renom, A.Rossi, F.Al-Shahrour, F.P.Davis, U.Pieper, J.Dopazo, A.Sali. "The AnnoLite and AnnoLyze Programs for Comparative Annotation of Protein Structures", in press.

185. Y. Cong, M. Topf, A. Sali, P. Matsudaira, M.T. Dougherty, W. Chiu, M.F. Schmid. "Crystallographic Conformers of Actin in a Biological Context" . Submitted.

186. M.A. Marti-Renom, U. Pieper, M.S Madhusudhan, A. Rossi, N. Eswar, F.P. Davis, F. Al-Shahrour, J. Dopazo, A. Sali. "DBAli tools: mining the protein structure space". *Nucleic Acid Research*, 2007.

187. F.P. Davis, D.T. Barkan, N. Eswar, J.H. McKerrow, A. Sali. "Potential Host–Pathogen Protein Interactions Predicted by Structure". Submitted.

188. C.R. Booth, A.S. Meyer, Y. Cong, M. Topf, A. Sali, S.J. Ludtke, W. Chiu, J. Frydman. "Mechanism of lid closure in the eukaryotic chaperonin TRiC/CCT." Submitted.

189. C.M. Anderson, D.L. Smith, S. Makovets, D. Korkin, A. Sali, E.H. Blackburn. "Tel2 Mediates Binding of Tel1 to Sites of DNA Damage." Submitted.

190. S. Dokudovskaya, L. Veenhoff, J.C. Kipper, W. Zhang, R. Williams, O. Karni-Schmidt, A. Suprapto, F. Alber, A. Sali, B.T. Chait, M.P. Rout. "Molecular Architecture of the Yeast Nuclear Pore Complex: A Comprehensive Proteomic Survey." Submitted.

191. F. Alber, S. Dokudovskaya, L. Veenhoff, D. Devos, B.T. Chait, M.P Rout, A.

Sali. "Molecular Architecture of the Yeast Nuclear Pore Complex: Satisfaction of Spatial Restraints Derived from Proteomic Data." Submitted

192. F. Alber, S. Dokudovskaya, L. Veenhoff, W.Zhang, J.Kipper, D. Devos, A.Suprapto, O.Karni, B.T. Chait, A. Sali, M.P Rout. " Molecular Architecture of the Yeast Nuclear Pore Complex: Satisfaction of Spatial Restraints Derived from Proteomic Data." Submitted

193. F. Melo, A. Sali. "Fold assessment for comparative protein structure modelling", submitted.

RESEARCH PROGRAM (SEPARATE SUMMARY)

The following five articles represent much of the research of the Sali group.

1. NEW METHODS FOR COMPARATIVE MODELING

B. John, A. Sali. "Comparative protein structure modeling by iterative alignment, model building, and model assessment" *Nucleic Acids Research* **31**, 1982-1992, 2003.

In the absence of an experimentally determined protein structure, homology or comparative protein structure modeling can often provide a useful 3D model of a given protein based on its similarity to a known structure. To increase the impact of comparative modeling on biology, we are maximizing its accuracy and applicability to as many proteins as possible. This is achieved by predicting and minimizing mistakes in fold assignment and sequence-structure alignment as well as errors in the modeling of sidechains, loops, and core regions. As any other prediction, the accuracy of comparative modeling is limited by the information that can be used in model calculations (*eg*, known related structures, energy functions). Therefore, to provide us with a flexible framework for continued development, we previously introduced comparative modeling by satisfaction of spatial restraints (Sali and Blundell, J. Mol. Biol. **234**, 779-815, 1993). This method computes models based on both a physical molecular mechanics force field as well as statistical preferences extracted from known protein structures and allows us to explore varied protein representations, scoring functions, and optimization protocols.

In the present paper, we developed and benchmarked an approach to minimize errors in the alignment of the target sequence against the template structure. Alignment errors are the single most important limitation on comparative modeling because they are frequent and have a large impact on the model accuracy. Unfortunately, no current comparative model building method can generally recover from errors in the input alignment. To ameliorate this problem, we developed an automated method that optimizes both the alignment and the model implied by it. This task is achieved within the framework of modeling by satisfaction of spatial restraints with the aid of a genetic algorithm protocol that starts with a set of initial alignments, and then iterates through re-alignment, model building, and model assessment to optimize a model assessment score. The average model accuracy increased from 43 to 54% (the model accuracy was measured

as the percentage of the C_α atoms of the model that were within 5Å of the corresponding C_α atoms in the superposed native structure).

An extrapolation indicates that our protocol might in principle be able to produce useful models for domains in ~110,000 protein sequences that could not be modeled previously with useful accuracy. Further development of comparative modeling is needed, as its impact on biology and drug discovery is assured to grow with the progress of structural biology.

2. COMPARATIVE MODELING TOOLS AND STRUCTURAL GENOMICS

U. Pieper, N. Eswar, H. Braberg, M.S. Madhusudhan, F. Davis, A.C. Stuart, N. Mirkovic, A. Rossi, M.A. Marti-Renom, A. Fiser, B. Webb, D. Greenblatt, C. Huang, T. Ferrin, A. Sali. "MODBASE, a database of annotated comparative protein structure models, and associated resources" *Nucl. Acids. Res.* **32**, D217-D222, 2004.

To increase the impact of comparative modeling on biology, we have also been maximizing its efficiency and accessibility to as many scientists as possible. We review here our integrated collection of programs, web servers, and databases that facilitate comparative modeling and contribute to the functional characterization of proteins based on structural considerations. These resources include a program for comparative modeling by satisfaction of spatial restraints (MODELLER), an automated software pipeline for large-scale comparative modeling (MODPIPE), a comprehensive database of comparative models for all protein sequences that are detectably related to at least one known protein structure (MODBASE), and several other programs and databases that contribute to functional annotation of the modeled sequences.

Currently, MODELLER is perhaps the single most widely used program for comparative protein structure prediction. It has been distributed to ~6,000 registered academic users as well as many pharmaceutical groups.

The genomic scale of biological data sets requires the automation of computational methods for their analysis and use. We automated the entire process of comparative modeling, including fold assignment, sequence-structure alignment, model building, and model assessment, and implemented it in the MODPIPE software, which relies on MODELLER.

We are continually applying MODPIPE to all known protein sequences. The resulting models are stored in MODBASE, which is linked bi-directionally with the SwissProt/TrEMBL (EBI, Cambridge, UK), PIR (Georgetown University), Protein Kinase Resource (UCSD), Genome Browser (UCSC), and GenMAPP (Gladstone Institute) databases.

Our large-scale comparative modeling results led us to suggest some key aspects of the current target selection strategy for the structural genomics initiative. Target selection is strongly dependent on the strengths and weaknesses of comparative modeling. We suggested defining targets such that most of the remaining sequences share at least 30% sequence identity to at least one known structure, thus ensuring largely correct alignments for comparative modeling.

3. PROTEIN FOLDING

C.M. Dobson, A. Sali, M. Karplus. "Protein folding: A perspective from theory and experiment" *Ang.Chem.Int.* **37**, 868-893, 1998.

As the number of possible conformations for a polypeptide chain is astronomically large, a systematic search for the native structure would require an almost infinite length of time. To solve this apparent paradox, the "old view" of protein folding suggested a simple sequential model, which postulates a unique folding pathway with defined and sequential intermediates.
In this article, we synthesized evidence from both experiment and theory about the so-called "new view" of protein folding, which suggests that neither the folding pathway nor the set of folding intermediates is unique. Instead, the folding process is heteregenous resulting from a complex interplay between entropic and enthalpic contributions to the free energy of the system during the reaction. This novel perspective had a strong impact on both experimentalists and theoreticians interested in the protein folding problem. The term "new view" was introduced in the News & Views article (R.L. Baldwin. Nature 369, 183, 1994) that accompanied our original paper on Monte Carlo simulations of a simple lattice model of protein folding (Sali *et al.*, *Nature* **369**, 248-251, 1994). In that publication, we determined the difference between folding and non-folding sequences for a simple 3D lattice model of a protein chain. Specifically, we generated many random sequences, tested them for their ability to fold to a stable native state, and finally compared the folding and non-folding sequences in terms of their structural and energetic properties. The necessary and sufficient condition for a sequence to fold rapidly was a pronounced energy minimum for the native state. This finding was explained by an elementary, three-stage random search mechanism of folding, which did not involve a pathway (thus the label "new view").

These Monte Carlo simulations, together with theoretical models and experimental results, provide the support for the current "funnel"-like view of the protein folding landscape. The understanding of folding is important for the analysis of many events involved in cellular regulation, the design of proteins with novel functions, the utilization of sequence information from the various genome projects, and the development of novel therapeutic strategies for treating or preventing human diseases that are associated with the failure of proteins to fold correctly.

4. APPLICATIONS OF BIOINFORMATICS TO SPECIFIC BIOLOGICAL SYSTEMS

D. Devos, S. Dokudovskaya, F. Alber, R. Williams, B.T. Chait, A. Sali, M.P. Rout. "Components of Coated Vesicles and Nuclear Pore Complexes Share a Common Molecular Architecture" *PLOS Biology* **2**, 1-9, 2004.

It is essential for developers of computational methods to collaborate with experimentalists in applying their methods to practical problems. Such applications provide a validation of the methods as well as feedback for further development. Over the last decade, we collaborated with numerous biologists, producing 33 papers focused on structural and functional characterization of 22 specific proteins and protein families (see *cv*), with 7 new collaborations in progress. We illustrate this effort by one specific application, an assessment of the predicted folds of

nucleoporins that indicated a possibility of a common evolutionary origin for coated vesicles and nuclear pore complexes.

Numerous features distinguish eukaryotes from prokaryotes, chief among which are the membrane systems that form elaborate compartments and vesicular trafficking pathways, and sequester the chromatin within the nuclear envelope. The nuclear pore complex is the portal that specifically mediates macromolecular trafficking across the nuclear envelope. Although it is generally understood that these internal membrane systems evolved from specialized invaginations of the prokaryotic plasma membrane, it is not clear how the nuclear pore complex evolved from organisms with no analogous transport system. We reported here a structural analysis of the seven proteins comprising the yNup84 subcomplex, a core building block of the nuclear pore. Our analysis revealed close similarities between the structures comprising the yNup84 subcomplex and those comprising the major types of vesicle coating complexes that maintain vesicular trafficking pathways. These similarities suggested a common evolutionary origin for the nuclear pore complex and coated vesicles in an early membrane-curving module that led to the formation of the internal membrane systems in modern eukaryotes. We hypothesized that the progenitor of the nuclear pore complex arose from a membrane-coating module wrapping extensions of an early ER around the cell's chromatin. This module, the yNup84 subcomplex, still serves to curve and stabilize the nuclear pore membrane in modern eukaryotes; as such, it functions as a key scaffold to form the nuclear pore complex, the portal of the nucleus.

This bioinformatics analysis illustrates how computational results can lead to illuminating hypotheses that rationalize the existing data and facilitate making new testable predictions.

5. A VIEW OF STRUCTURAL BIOLOGY
A. Sali, R. Glaeser, T. Earnest, W. Baumeister. "From words to literature in structural proteomics" *Nature* **422**, 216-225, 2003.

In this article, we presented a view of structural biology that guides our development and application of a computational system for enumerating structures of protein assemblies that are consistent with all available information from experimental methods, physical theories, and statistical preferences extracted from biological databases. To achieve this objective, we are extending the formalism of modeling by satisfaction of spatial restraints from individual proteins to assemblies of proteins.

Technical advances on several frontiers have expanded the applicability of existing methods in structural biology and helped close the resolution gaps between them. As a result, we are now poised to integrate structural information gathered at multiple levels of the biological hierarchy — from atoms to cells — into a common framework. The goal is a comprehensive description of the multitude of interactions between molecular entities, which in turn is a prerequisite for the discovery of general structural principles that underlie all cellular processes. In contrast to structure determination of individual proteins, structural characterization of macromolecular assemblies usually requires diverse sources of information. This information may vary greatly in

terms of its accuracy and resolution, and includes data from both experimental and computational methods, such as X-ray crystallography, NMR spectroscopy, electron microscopy, chemical cross-linking, affinity purification, yeast two-hybrid system experiments, calorimetry, computational docking, and bioinformatics analysis of protein sequences and structures. New computational methods are needed for computing the 3D models of a given protein assembly that are consistent with all available information about its composition and structure.

Early versions of this generalized approach were used to contribute to the structure determinations of the first eukaryotic ribosome from yeast (in collaboration with Joachim Frank and Gűnther Blobel: Spahn et al. Cell 107, 373, 2001), the *E. coli* ribosome (with Joachim Frank: Gao et al. Cell 113, 789, 2003) as well as to structure characterization of the yeast nuclear pore complex (with Mike Rout and Brian Chait, unpublished). In all of these cases, our integrated system improved efficiency, accuracy, resolution, and completeness of the structural coverage of the studied assemblies.